Journal of

CHRONIC DISEASES

VOL. 4 SEPTEMBER, 1956 NO. 3

VOL.
4
956

Original Communications

ONSET AND EARLY COURSE OF ESSENTIAL HYPERTENSION

RUTH B. BENEDICT, M.D.

WASHINGTON, D. C.

From the Department of Medicine, George Washington University School of Medicine, and The Group Health Association, Inc.

(Received for publication Feb. 25, 1956.)

ESSENTIAL hypertension is a common disturbance. The literature on the subject is vast. Nonetheless, most of our knowledge of the onset of the disorder is based on deduction rather than on clinical observation. Fishberg,[1] in his recently revised text, "Hypertension and Nephritis," confines himself to speculative remarks on the early phase of the disease. The "Bibliography on the World Literature on Blood Pressure 1920-1950"[2] contains no entry which bears directly on the subject. My own search of the literature, while not exhaustive, has revealed one set of papers,[3-6] which deals factually with the problem. This paucity of information is presumably because few patients are under continuous medical observation by one doctor or clinic both before and after the onset of high blood pressure. Since the start of the disorder often cannot be accurately dated, knowledge of its early course is equally incomplete. The opportunity to study both these aspects of hypertension presented itself in Washington, D. C., at the Group Health Association, where many individuals have received continuous medical care for over a decade. This paper discusses the onset and early stage of hypertension as it was found in persons who developed it while members of the Association during the last thirteen to seventeen years. Tabulations of various aspects of hypertensive disease will be presented, and such conclusions as appear to be warranted will be discussed.

The population from which these patients were drawn consisted of about 3,500 predominantly healthy, white, middle class Federal workers and their families. The majority were in the middle echelons of government service and were well above average in intelligence and education. Roughly one-third were in their twenties, one-third in their thirties, and one-fifth in their forties when first seen. All were over 21 years of age. Approximately 54 per cent were women and 46 per cent were men. In reviewing their charts, I was able to find 149 originally normotensive persons who later developed a diastolic pressure of 90 mm. Hg or more, usually accompanied by a systolic pressure of 150 mm. Hg or more. Not included in the 149 were individuals who went on to develop only systolic hypertension of 150 mm. Hg or more. Most of these latter were elderly people whose increased pressure was a reflection of arteriosclerosis of the aorta and large arteries. A few were young nervous persons who reacted to medical examination with rapid heart rates and wide pulse pressures. Forty-three of the original 149 patients were rejected from the study. In some, the blood pressure was elevated only sporadically. The hypertension of others appeared to be on the basis of kidney disease other than chronic renal infection. In still others, less than three early normal blood pressure readings had been recorded, or the interval between the last normal and the first high reading was greater than four years. This left a total of 106 patients for investigation.

On the whole, the records were adequate for the nature of the study. The number of blood pressure readings on each chart varied from 8 to more than 50. The mean number of readings was 17. The time interval between readings varied from a few days to several years. In all but a few instances a minimum of one reading for each 2 years was recorded in the normotensive period, while one or more readings per year were generally recorded after hypertension developed. Fifteen charts contained no electrocardiograms and 9 no chest x-rays. Such omissions were usually found in the records of young persons with slight hypertension in whom these procedures would probably have revealed normal results. A more serious gap in many charts was the absence of periodic examinations of the optic fundi, so that it was impossible to investigate this aspect of hypertensive disease.

Each patient had been followed for 1.5 to 15 years after the onset of high blood pressure. Three-fourths had been observed for five years or longer. Only a handful had received modern antihypertensive therapy toward the end of the follow-up period. In the few cases where such treatment appeared to be effective, the period after which it had been started was omitted from the investigation. It was therefore possible to study the early course of untreated hypertension. As might be expected in an investigation based on white office patients, it was found that most subjects did not develop a serious form of the disorder during the period of observation. As a result, this study is primarily a portrayal of the onset and early course of mild hypertension which, as is well known, comprises the bulk of all hypertensive disease.

AGE OF ONSET OF HIGH BLOOD PRESSURE

The date of onset of hypertension was taken to be the beginning of that period when at least 3 of 5 consecutive blood pressure readings were elevated and after which all but an occasional reading were high.

Table I presents the age of onset by decades. There was no significant difference therein between the sexes. The average age of onset in all patients was somewhat less than 44 years.

TABLE I. AGE OF ONSET OF HYPERTENSION

AGE	MEN	WOMEN	BOTH
21–29	2	3	5 = 5%
30–39	18	17	35 = 33%
40–49	18	22	40 = 38%
50–59	9	14	23 = 22%
60–69	2	1	3 = 2%
Total	49	57	106

OL.
4
956

HEIGHT OF THE BLOOD PRESSURE AT ONSET

The height of the mean diastolic pressure at onset was in the 90's in 75 per cent of the patients and in the low 100's in the remainder.

The charts of 53 patients contained 5 or more blood pressure readings taken before the start of established hypertension. At least one elevated reading in the prehypertensive period was found in almost 80 per cent (42 patients). Nonetheless, in most persons the onset of permanent hypertension was reasonably clear-cut and fell within a period of several years. In 8 cases, with a more gradual onset, there was a span of four to eleven years in which frequent elevated readings were recorded, although the mean pressure for the entire period was still normal.

SYMPTOMS NEAR ONSET OF HIGH BLOOD PRESSURE

A study was made of the reasons given by patients for seeking medical attention at the time when high blood pressure was first observed. Headache, nonspecific chest pain, palpitation, dizziness, epistaxis, and manifestations of cerebral or retinal vasospasm were listed as "hypertensive" symptoms. There was no substantial difference in such complaints between the sexes. Seventy-eight per cent of the patients were asymptomatic early in the disorder. When symptoms occurred, they were relatively more frequent in those with diastolic pressure in the 100's.

Eleven of the 23 patients who mentioned "hypertensive" symptoms near onset had already complained of similar disturbances while normotensive. In this connection it is of interest that one of the two patients with symptoms and ophthalmoscopic evidence of retinal vasospasm had presented similar difficulties for many years before the start of hypertension.

FACTORS OF POSSIBLE ETIOLOGIC SIGNIFICANCE IN HIGH BLOOD PRESSURE

The relationship between the onset of hypertension and the occurrence of the menopause was studied. Those women whose hypertension started two

or more years before the cessation of menses were placed in the premenopausal group. Those in whom it started five or more years after the termination of menses or the onset of flushes were put in the postmenopausal group. The remainder were placed in the menopausal group. Table II shows that hypertension and the menopause occurred concomittantly in about one-third of the women.

TABLE II. ONSET OF HYPERTENSION IN RELATION TO THE MENOPAUSE IN 57 WOMEN

Premenopause	26 = 46%
At menopause	17 = 30%
Post-menopause	12 = 21%
Cannot tell	2 = 3%

Although many of these women had had children, only 17 were observed during one or more pregnancies. In one individual, mild hypertension started during pregnancy. Her chart showed 1 high and 3 normal readings in the ten months preceding the pregnancy. In 5 women, hypertension first occurred within two years after a pregnancy, and in 3 of them an occasional elevated reading had been recorded during the preceding pregnancy. No instance of toxemia of pregnancy followed by permanent hypertension was observed.

Twenty-four patients displayed various urologic lesions at some time before the start of hypertension. There was evidence, as judged by symptoms, urine examinations, urine cultures, pyelograms, and blood chemistries, that chronic renal infection has persisted and possibly played a causative role in the high blood pressure of only one of them. In this man, a mean blood pressure of 142/100 mm. Hg was associated with chronic pyelonephritis and the eventual development of uremia.

TABLE III. INCIDENCE OF OVERT NERVOUSNESS IN PATIENT AND CONTROLS

	MEN		WOMEN		BOTH	
	PATIENTS	CONTROLS	PATIENTS	CONTROLS	PATIENTS	CONTROLS
Nervous	28 = 57%	12 = 25%	45 = 79%	21 = 37%	73 = 69%	33 = 31%
Not nervous	21 = 43%	37 = 75%	12 = 21%	36 = 63%	33 = 31%	73 = 69%
Total	49	49	57	57	106	106

The incidence of emotional disturbance was studied in the patients and in a control group of 106 individuals matched for age and sex and drawn at random from members of the original population. Persons were placed in the disturbed category if they had often complained of tension, if the doctors who treated them had noted over the years that they appeared to be nervous, if the consulting psychiatrist who saw some of them had diagnosed a psychoneurosis, or if they had embarked upon a course of psychotherapy. Persons who complained of

tension only sporadically or who had frequent psychosomatic difficulties but made no mention of nervous factors, were classified as emotionally undisturbed, since no adequate evaluation of their symptoms could be made. The disturbed group therefore consists of "overtly nervous" patients and fails to include others who may have been equally, but less obviously, affected. Table III shows that more than two-thirds of the patients and less than one-third of the controls were "overtly nervous." In both patients and controls, there was a higher incidence of "nervousness" in women. Further analysis of the patients' charts revealed that 80 per cent of those showing emotional difficulties after the onset of high blood pressure had displayed similar problems while normotensive.

EARLY COURSE OF HYPERTENSION

Eighty patients were observed for at least five years after the start of high blood pressure. In these, a comparison was made between the height of the mean diastolic pressure in the first and second halves of the initial five years of hypertension. Table IV shows that the diastolic pressure remained stable during the first five years in 84 per cent and increased by 10 mm. Hg or more in the remainder. By the close of the study, 12 patients, followed from 7 to 14 years after the beginning of hypertension, displayed diastolic pressures of 110 mm. Hg or more.

TABLE IV. COMPARATIVE HEIGHT OF DIASTOLIC PRESSURE FIVE YEARS AFTER ONSET OF HYPERTENSION IN 80 PATIENTS

	MEN	WOMEN	BOTH
Remained at same level	29	38	67 = 84%
Increased 10 mm. Hg or more	8	5	13 = 16%
Total	37	43	80

HEADACHES IN EARLY HYPERTENSION

Recurrent headache was the most frequent symptom seen in early hypertension. This complaint also occurred in a substantial number of the normotensive controls as shown in Table V. It was present in 36 per cent of the patients and 19 per cent of the controls. In each group headaches were twice as frequent in women as in men. Nervous factors were frequently mentioned in the charts of most headache patients, and a diagnosis of tension headache had been made in over one-half. Among the hypertensives, headaches were relatively more frequent in those with higher diastolic pressures. Twenty-eight of the 38 hypertensives with headaches had already complained of cephalalgia while normotensive. In 6 of these, headaches appeared to become worse after the start of high blood pressure. On the other hand, there were also 6 persons who had suffered from recurring cephalalgia while normotensive and who stopped complaining thereof after the development of hypertension.

TABLE V. INCIDENCE OF HEADACHES IN HYPERTENSIVE PATIENTS AND CONTROLS

	MEN		WOMEN		BOTH	
	PATIENTS	CONTROLS	PATIENTS	CONTROLS	PATIENTS	CONTROLS
Headaches	12 = 24%	6 = 12%	26 = 46%	14 = 24%	38 = 36%	20 = 19%
No headaches	37 = 76%	43 = 88%	31 = 54%	43 = 76%	68 = 64%	86 = 81%
Total	49	49	57	57	106	106

VASCULAR DISEASE IN EARLY HYPERTENSION

Cardiovascular lesions occurring early in the course of hypertension were studied. They consisted of arteriosclerotic and hypertensive heart disease, arteriolonephrosclerosis, and cerebral and retinal artery thromboses. The diagnosis of asymptomatic heart disease was made on the basis of x-ray evidence of left ventricular enlargement, or on the electrocardiographic changes of left ventricular hypertrophy and strain, or bundle branch block, or on a combination of x-ray and electrocardiographic abnormalities. Symptomatic coronary artery disease was diagnosed by the accepted criteria for angina pectoris, coronary insufficiency, and myocardial infarction. Table VI shows the occurrence of these lesions in patients and controls. Not included in the table are 2 female and 1 male hypertensives whose electrocardiograms showed ST-T wave changes compatible with early left ventricular strain.

TABLE VI. TYPE OF VASCULAR DISEASE ENCOUNTERED IN PATIENTS AND CONTROLS

	MEN		WOMEN	
	PATIENTS	CONTROLS	PATIENTS	CONTROLS
Angina, coronary insufficiency, or infarct	8	3	5	0
Congestive heart failure	2	0	1	0
Asymptomatic hypertensive and/or arteriosclerotic heart disease	5	2	2	0
Cerebral or retinal artery thrombosis	3	1	2	0
Nephrosclerosis	1	0	1	0

More than one complication in 5 male and 3 female hypertensives and in 1 control.

The incidence of cardiovascular complications in male and female hypertensives and controls is presented graphically in Fig. 1, where the per cent afflicted is plotted as a function of time after the onset of hypertension. The base time for individuals in the nonhypertensive groups was taken as a time after the date of their initial observation equal to the average interval from initial observation to onset of hypertension in those of the same sex who developed the disease. This interval was 6.6 years for the males and 7.2 years for the females. The incidence was calculated by the life-table method described by Merrell and

Shulman.[7] The upper portion of Fig. 1 plots the percentage of patients remaining in the samples at yearly intervals. The four groups are seen to be quite comparable in this respect. Although some patients were observed for longer than the eight years covered in Fig. 1 and developed complications in the subsequent period, it is believed that the size of the remaining samples does not warrant further projection. It will be noted that in four males the onset of cardiovascular disease (in the form of coronary artery disease) antedated the onset of hypertension by 1, 4, 5, and 9 years. However, the graph shows that the subsequent rate of development of complications was markedly greater in the patients than in the controls and in the males than in the females.

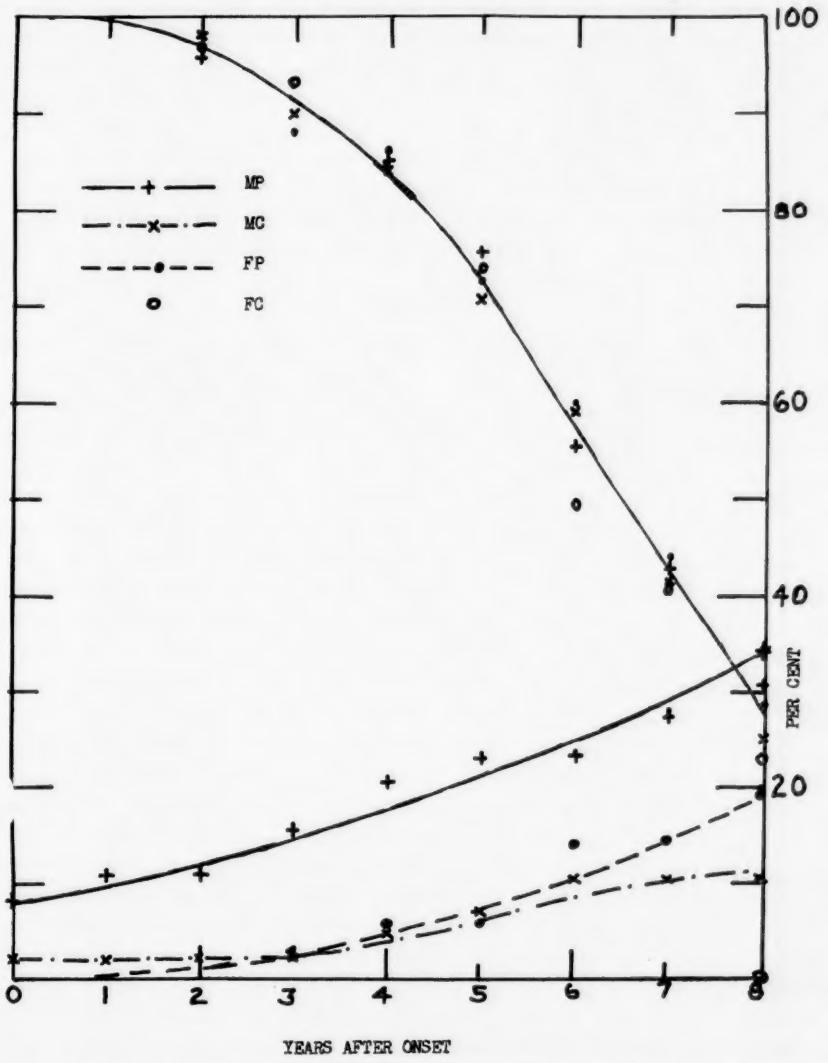

Fig. 1.—Incidence of cardiovascular complications in patients and controls. MP = male patients; MC = male controls; FP = female patients; FC = female controls. (No cases in the last group.) The points and curve in the upper part of the graph give the percentage of the original groups remaining in the samples at yearly intervals.

Complications were more frequent in the older patients as shown by the fact that the average age of onset of high blood pressure was 48 years in those who developed cardiovascular lesions and 42 years in those who did not.

The height of the diastolic pressure exerted an even greater effect on the appearance of cardiovascular difficulties. All of the 5 men and 3 of the 7 women whose diastolic pressures rose to 110 mm. Hg, or more, developed such lesions. Three of these developed both cardiac and cerebral difficulties, while only 2 of the 14 having complications with lower pressures eventually exhibited disease in more than one system. In addition, those with higher pressures were 1.5 years younger than average at the onset of hypertension while those with lower pressures and complications were 4.5 years older.

An interesting difference relating to symptomatic coronary artery disease occurred between the sexes. This lesion, like all others except nephrosclerosis, was more common in males. It first manifested itself at an average age of 50 in the men and 56 in the women. One-half of the men and only one woman were under 50 at its onset. One-half of the males already had symptoms of the disorder while normotensive, and in 3 of the remaining 4, symptoms appeared within the first four years of high blood pressure. On the other hand, all but one woman had been hypertensive for at least six years before coronary disease manifested itself.

DISCUSSION

In the patients here studied, the majority first developed high blood pressure in their thirties and forties. The average age of onset was somewhat less than 44 years. However, this should be less for all persons with essential hypertension since two-thirds of the original population from which these patients were drawn was already over 29 and since it included no one under 21. In his latest report, which deals with 150 subjects followed from the start of hypertension to death, Perera[6] found the mean age of onset of hypertension to occur in the early thirties. The earlier onset which he reports is apparently due to a number of factors. His sample is different from, and perhaps less representative than, the present one, since it was drawn from a hypertensive clinic and not from a homogeneous group of individuals most of whom were healthy when first observed. He was able to include a relatively large number of persons whose hypertension began before the age of 30. About one-third of his patients were Negroes, and it is probable that high blood pressure begins earlier in an appreciable number of that race.[8,9] He excluded patients over 50 with diastolic pressures in the 90's since he felt that in older people such pressures are attributable to arteriosclerosis of the large vessels. While sclerosis of the aorta may be responsible for this type of hypertension in some older persons, there are others in whom peripheral vasoconstriction also seems to play a role. As noted by Page,[10] arteriosclerosis and essential hypertension are often so inextricably intertwined that it is impossible to dissociate them. There were older individuals in this series whose pressure crept up gradually over the years and in whom this phenomenon may have been secondary to aortic sclerosis. On the other hand, there were patients in their fifties and sixties who had repeated diastolic pressures in the 70's for many years, and who then showed a rise of over 20 mm. Hg in this reading in less than a year's time. Furthermore, it is difficult to ascribe diastolic pressures

in the 90's to arteriosclerosis alone in patients who continue to run systolic pressures below 150 mm. Hg, since hypertension secondary to large vessel sclerosis is usually predominantly systolic. Six patients in this study developed this type of blood pressure after the age of 50. Perera also encountered no patient in whom essential hypertension started after the age of 48. In all of his subjects who developed diastolic pressures above 100 mm. Hg after that age, the hypertension could be ascribed to other causes. Such was not the case in this study, which includes nine patients who were normotensive till their fifties and sixties, and in whom no cause other than essential hypertension could be found to account for the eventual development of hypertension with diastolic pressures over 100 mm. Hg. It is therefore my impression that the mean age of onset of essential hypertension in general probably falls somewhere in the late thirties.

Both Perera's[6] study and the present one demonstrate that the onset of hypertension is most often asymptomatic. Most of the "hypertensive" symptoms which were noted here at onset were nonspecific and of a type frequently experienced by normotensive nervous individuals. Symptoms of retinal vasospasm are more closely related to the basic hypertensive process. Seemingly they may not only accompany the start of the disorder, but may occasionally be present for many years before. Hines[11] made a similar observation when he noted that certain individuals displayed fundoscopic evidence of arterial spasm years before they developed high blood pressure.

Although a third of the women first developed hypertension at the menopause, there seemed to be no causal relationship between the two, since the age of onset of high blood pressure was the same in both sexes. Corcoran and Page[12] reached the same conclusion on observing that castrated women were no more prone to hypertension than women in the general population.

"Overt nervousness" was present in over two-thirds of the patients and in less than one-third of the controls. In the majority of patients, nervousness predated the hypertension. This supports the belief that long-standing emotional disturbances are frequent in hypertensives and may play some role in the etiology of the disorder.[13,14]

Analysis of the headaches experienced by patients and controls points to the probability that many persons in both groups were chronic sufferers from tension headaches. Such a conclusion is compatible with the findings that both headaches and nervousness were more frequent in patients than in controls, that women (who are known to be more subject to tension headaches)[15] were more often afflicted than men, that nervous factors were prominently mentioned in the charts of most headache victims, and that, in the patients, headaches often predated hypertension and frequently became no worse after its onset.

The deleterious effect of hypertension on the vascular system is dramatically portrayed in the graph comparing the development of cardiovascular lesions in patients and controls. Specific factors influencing the appearance of these lesions were sex, age, duration and height of blood pressure, and the presence of previously existing or concomitantly developing arteriosclerosis. Since men are more prone to coronary (and other) arteriosclerosis at an earlier age, and since high blood pressure exerts an unfavorable effect on this process, it is not surprising that the hypertension carries a more serious prognosis in males.

SUMMARY AND CONCLUSIONS

1. The average age of onset of hypertension in 106 patients was somewhat less than 44 years, but the mean age of onset of all essential hypertension is probably lower.

2. In the majority in whom this factor could be studied, the onset of permanent hypertension fell within a period of several years.

3. The mean diastolic pressure at onset was in the 90's or low 100's in all.

4. The majority of patients were asymptomatic early in the disorder.

5. "Overt nervousness" was more than twice as frequent in patients as in controls. It occurred in two-thirds of the patients and usually preceded the hypertension.

6. About one-third of the women developed high blood pressure during the menopause, but the relationship between the menopause and hypertension did not appear to be causal.

7. Eighty-four per cent of 80 patients maintained the same range of diastolic pressure during the first five years of their disorder.

8. Headaches were the commonest complaint early in the disorder. Nervous tension seemed to play an important role in their production.

9. Vascular lesions were more frequent in patients than in the controls, and in hypertensive men than in hypertensive women. Age, duration and height of blood pressure, and the presence of pre-existing or concomitantly developing arteriosclerosis were important factors in the appearance of cardiovascular lesions.

I am indebted to Mr. William Haenszel, Head of the Biometry Section, National Cancer Institute, Bethesda, Maryland, for statistical guidance.

REFERENCES

1. Fishberg, A. M.: Hypertension and Nephritis, Philadelphia, 1954, Lea & Febiger, pp. 693-694, 758-761.
2. Kolter, E. K., and Katz, J. Ed.: A Bibliography of the World Literature on Blood Pressure 1920-1950, USA, 1952, Commonwealth of Massachusetts.
3. Perera, G. A.: Diagnosis and Natural History of Hypertensive Vascular Disease, Am. J. Med. **4**:416, 1948.
4. Perera, G. A.: The Life History of 100 Patients With Hypertensive Vascular Disease, Am. Heart J. **92**:421, 1951.
5. Perera, G. A.: The Natural History of Hypertensive Vascular Disease. In Hypertension, A Symposium, edited by E. T. Bell, Minneapolis, 1951, University of Minnesota Press, pp. 363-369.
6. Perera, G. A.: Hypertensive Vascular Disease; Description and Natural History, J. CHRON. DIS. **1**:33, 1955.
7. Merrell, M., and Shulman, L. E.: Determination of Prognosis in Chronic Disease, Illustrated by Systemic Lupus Erythematosis, J. CHRON. DIS. **1**:12, 1955.
8. Weiss, M. M., and Prussmack, J.: Essential Hyptertension in the Negro, Am. J. M. Sc. **213**:475, 1938.
9. Hunter, W. S.: Coronary Occlusion in Negroes, J. A. M. A. **131**:12, 1946.
10. Page, I. H., in a discussion of Ref. No. 5 in Hypertension, A Symposium, edited by E. T. Bell, Minneapolis, 1951, University of Minnesota Press, p. 370.
11. Hines, E. A.: The Significance of Vascular Reaction As Measured by the Cold Pressor Test, Am. Heart J. **19**:408, 1940.
12. Corcoran, A. C., and Page, I. H.: Menopausal Hypertension, a Critical Study, Am. J. M. Sc. **213**:475, 1947.
13. Schroeder, H. A.: Pathogenesis of Hypertension, Am. J. Med. **10**:189, 1951.
14. Benedict, R. B.: Psychiatric Concepts in Essential Hypertension, M. Ann. District of Columbia **23**:197, 1954.
15. Friedman, A. P., deSola Pool, N., and Von Storch, T. J. C.: Tension Headache, J. A. M. A. **151**:174, 1953.

OL.
4
956

STUDIES IN HYPERTENSION

I. An Epidemiologic Approach to the Study of the Natural History of Essential Hypertension

E. Gurney Clark, M.D., Dr. P. H., Charles Y. Glock, Ph.D., and Robert L. Vought, M.D., M.P.H.

New York, N. Y.

From the Division of Epidemiology, School of Public Health and Administrative Medicine of the Faculty of Medicine, and the Bureau of Applied Social Research, both of Columbia University

(Received for publication May 5, 1956.)

THIS paper is a progress report of a five-year program* of planning and research in the epidemiology of essential hypertension. Its purpose is to outline the development, accomplishments, and future plans of this program. The development of the program illustrates how an epidemiologic method[1] is applied to a study of the natural history of essential hypertension. The accomplishments to date are shown in part by a series of papers to follow the present one.[2-6] The future plans are indicated by a discussion herein of the important elements of a long-term epidemiologic field study of the natural history or evolution of this disorder.

Clinical and laboratory studies of hypertension, concentrated primarily on the period of pathogenesis or the natural course of the disease in man, have resulted in the development of therapeutic measures which alter or delay the progress of signs and symptoms. However, they have not yielded information that will satisfactorily prevent this disorder. This is because so little is known about influences which actually initiate pathogenesis. From the viewpoint of epidemiology and disease prevention, the concept of natural history of any disease is not limited to its natural course in an individual, but also includes known and suspected agent, host, and environmental elements that might influence onset.[1,7] This concept of natural history, first suggested by Perkins,[8]

Other members of the project staff have been: John A. Morsell, Ph.D.; Jacob Katz, M.A.; Ernest K. Koller, B.A.; Henry L. Lennard, Ph.D.; Morton D. Schweitzer, Ph.D.

*Supported since 1950 by funds from a Recess Commission on Hypertension. The Resolves of 1949 of the General Court (Legislature) of Massachusetts created this Commission upon the recommendation of Representative Paul A. McCarthy. This Commission, under the Chairmanship of Dr. Reginald H. Smithwick, had the stated purposes of fostering (a) studies of the natural history of hypertension; (b) clinical investigations of the disorder fully developed in man, with an appraisal of the various forms of therapy now available to combat it; and (c) basic research on its pathologic physiology in man and in the experimental animal. In pursuit of the first of these purposes, the Commission enlisted the collaboration of the Division of Epidemiology, which called upon the Bureau of Applied Social Research to serve as a partner.

is illustrated schematically for hypertension in Fig. 1. The "twilight zone" of onset,[9] which precedes the course of the disorder, is in turn preceded by the interaction of some of these or other presently unknown etiologic factors. The principal factors which have been suggested as bearing some relation to the onset of hypertension are shown in the prepathogenesis period of Fig. 1.

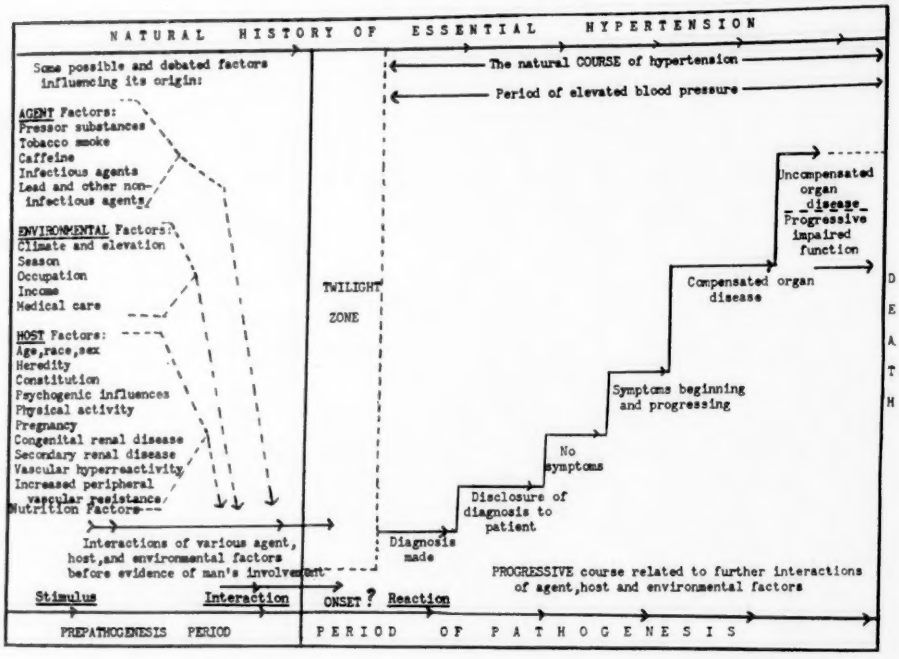

Fig. 1.—The natural history of essential hypertension. (Adapted from data presented at "A Symposium on Essential Hypertension.")[14]

Studies of the relative importance of the etiologic elements operating during the prepathogenesis and "twilight" phases should not be limited to clinical and laboratory investigation. Epidemiologic field studies should be undertaken in the living and working environment of hypertensives and nonhypertensives over a period of time, in order to determine the degree of association between levels of blood pressure on the one hand and the variety of personal and environmental factors thought to be of possible etiologic significance on the other. This has been the familiar and successful experience in studies of infectious disease and should be equally helpful in essential hypertension.

The *ultimate* objective of such a program is the establishment of a long-term field study of blood pressure in one or more communities. Repeated observations should be made on representative samples of the population in order to collect data on the range of variables now thought to be of possible significance in relation to the onset and development of hypertension.

The design of a field study of this kind requires careful appraisal of the known and suspected etiologic factors and of the methods of their qualitative and quantitative measurement. The plan adopted for a systematic epidemiologic

approach conforms with five steps already briefly outlined.[1] The present paper summarizes the procedure followed in the first three of these steps and outlines the plans for steps *IV* and *V*.

I. A More Precise Definition of the Nature, Extent, and Significance of the Problem and a Clarification of the Objectives; II. A Critical Appraisal of Existing Information.—In this organized epidemiologic approach these two steps are required for an evaluation of the need and feasibility of conducting epidemiologic inquiries into the natural history of hypertension in man. Several interrelated activities were therefore initiated: a review of the literature on incidence, prevalence, trend, and other epidemiologic features of hypertension was made; a marginal punch-card cross reference file of the world literature on blood pressure was made for staff use*; plans were formulated for organizing and conducting a symposium on the epidemiology of hypertension similar to one which had been held a short time before on the Epidemiology of Mental Disorders.[10]

The review of the literature on incidence and prevalence completed in 1950[11] revealed the following: (a) there are no consistent standards of blood pressure measurement; (b) there are no uniform criteria for defining hypertension; (c) none of the existing studies is based on a representative sample of the population of a country or suitable subdivision of it; and (d) although various studies leave no doubt that hypertension is a health problem of great magnitude which probably varies according to age, sex and race, they do not answer the question as to its exact extent, nor provide an accurate account of its distribution in any population.

The immediate objective of this program of planning and research in hypertension is the formulation of a practical design for a long-term field study of blood pressure. Through such a study it is hoped ultimately (a) to develop criteria to aid in the diagnosis of essential hypertension; (b) to identify and evaluate the factors which might initiate this disorder; and (c) to provide information which might lead to the interruption of its natural evolution before serious illness develops.

The initial review on incidence and prevalence and the cross reference file, both completed during 1950, were helpful in the organization of the Symposium, planned to contribute to the appraisal of existing information. The Symposium† (held in 1951) was attended by an interdisciplinary group of 36 scientists selected to include leading workers who were interested in hypertension.[13] It provided a comprehensive review of what was then known about agent, host, and environmental factors in the evolution of essential hypertension and furnished counsel as to a study or series of studies on its evolution.

The Symposium group enumerated problems to which field studies might provide important information as follows: a definition of hypertensive disease; the prevalence of elevated blood pressure according to such categories as age, race, and sex; the natural history of hypertensive disease in man; the significance

*This file was published in 1952 in three volumes[12]: an author index; a classified list of 16,460 articles on blood pressure (1920-1950); and a volume of 1,400 abstracts on certain epidemiologic aspects of hypertension.

†The proceedings of this Symposium were edited by the project staff and published in book form by the Commonwealth of Massachusetts.[14]

of blood pressure levels in relation to prognosis; the phenomenon of hyperreactivity; the characteristics of the prehypertensive period; the influence of heredity and constitutional factors; blood pressure levels in relation to pregnancy, injury, infection, occupation, growth, habits, and mode of life; and the role of major psychologic trauma in the development of hypertension.*

In view of the complexities of the hypertension problem and of the requirements of scientific design of a long-term field study, the Symposium suggested that a preliminary period of planning and experimentation be undertaken in order to develop specific techniques for field use. The literature on hypertension is replete with reports of studies which suggest that this or that factor is a decisive element in producing hypertension. While few of these studies have provided conclusions which have been completely validated, many of them present hypotheses which should be subjected to tests among both hypertensive and nonhypertensive subjects. Therefore, it was necessary to select for further testing those hypotheses which give most promise of adding to knowledge about the natural history of this disorder; and to determine the most efficient techniques for testing them in a long-term study.

Critical appraisals of the literature were therefore undertaken of the following subjects in relation to hypertension: blood pressure measurement, the phenomenon of vascular hyperreactivity, demographic and social characteristics, psychologic and psychiatric factors, genetic and hereditary factors, constitutional factors, nutritional factors, and physiologic and medical factors. The review of the literature of five of these subjects have been completed. Of these five subjects, the first—blood pressure measurement—is considered in the next two papers of the present series[2,3]; the next—vascular hyperreactivity—is dealt with in the fourth paper[4]; and in each of these papers there has been added personal field observations.

Critical analyses of the literature as to psychologic and psychiatric factors[5] and as to certain demographic and social characteristics[6] also appear in this series of papers, but in each instance, as yet, without our own data from field study.

Each reviewed article has been studied from the standpoint of scientific methodology, the soundness of the research design, the validity of the analyses of results, and the extent to which results can be utilized outside the experiment itself.

Appraisal of information on blood pressure measurement: A long-term study of hypertension requires that some procedure be established to assure the detection of significant changes in the level and variability of blood pressure at intervals over long periods of time. Two kinds of problems are involved: one, the accuracy of measurement at any one time, i.e., the relationship of sphygmomanometer values and the true values of an intra-arterial manometer; and two, the representativeness of measurements for purposes of comparing the status of peripheral vascular resistance at successive points in time.

The resolution of these two problems and of the general problem of blood pressure measurements for meeting the needs of a long-term study requires answers to three basic questions: (a) Under what conditions should blood

*Page 325, reference 14.

pressure measurements be made? (b) How many readings at any one time are necessary to obtain a representative picture of peripheral vascular resistance at that time? (c) What should be the length of time between observations in a long-term study? These questions are dealt with in papers II and III of this series.[2,3]

III. Formulation of Hypotheses and the Testing of Procedures To Be Used in the Study Design.—The determination of data to be collected and the methods of collection involves two fundamental decisions: (1) what has to be known about an individual aside from blood pressure measurements to understand the factors which might contribute to or explain the onset and development of hypertension or its failure to occur; and (2) how this knowledge can be obtained within the framework of a long-term study. In order to make these decisions it is necessary to review the many hypotheses about the relationship between different variables and the etiology of hypertension and to choose those which can be tested satisfactorily in a sample of the population over a period of time. In some instances it is necessary to develop techniques with which these hypotheses might effectively be tested. It is important that, in keeping with scientific methodology, the formulation of hypotheses should be carried as far as possible before a long-term study is actually begun. Since other hypotheses will perhaps arise during the course of the study, the design must be sufficiently flexible to allow for their inclusion in the testing process of the field study.

After reviewing and evaluating many of the factors hypothesized as having some association with hypertension, some hypotheses have been formulated which can be incorporated for testing in a long-term prospective study, and some which can be tested concurrently and retrospectively. The categories in which hypotheses have been explored in this project and those in which further work is planned are presented below.

Hypotheses concerning the phenomenon of vascular hyperreactivity: Several questions had to be answered before a decision could be reached as to the inclusion of tests of hyperreactivity in a long-term study. Is the hypothesis that hyperreactivity is associated with the subsequent development of hypertension on sufficiently sound grounds scientifically to warrant further exploration in a long-term study? Are there specific, consistent, and reproducible methods for defining, recognizing, and measuring hyperreactivity? Are any of the existing tests suitable for use under field conditions?

The fourth paper of this series reports on the efforts to secure answers to these questions by a small scale epidemiologic study[4] conducted in Boston in 1953. Although the field study provided definitive answers to only one of the specific questions above (i.e., the tests *are* adaptable to field conditions), it became evident that the phenomenon of vascular hyperreactivity should be taken into account in a long-term study. However, further work of a methodologic nature must be undertaken before a decision can be reached as to *how* to include measurement of this phenomenon in a study in order to test the hypotheses that this phenomenon is a precursor of essential hypertension.

Psychologic and psychiatric hypotheses: It is difficult to isolate the hypotheses regarding the relationship of these factors to essential hypertension, although much work has been done in this area. The current hypotheses about psychologic

factors associated with the onset or presence of hypertension are examined in the fifth paper of this series.[5] Three approaches to the nature of these psychologic factors are recognized. The first includes those hypotheses emphasizing the whole configuration of characteristics making up the so-called "hypertensive personality." The second includes statements considering specific psychologic conflicts as the unit of psychologic description in relation to hypertension. The third encompasses a variety of hypotheses dealing with stressful events. Further evaluation of the observations which led to these hypotheses must be made. Furthermore, research procedures for testing them in a long-term study must also be developed.

Hypotheses concerning demographic and social characteristics: The demographic and social conditions under which disease is likely to develop or not to develop can be examined more effectively by epidemiologic than by clinical or laboratory procedures, but all three are essential to a full understanding of natural history. The following hypotheses can be tested by epidemiologic studies: (1) the occurrence of hypertension, its mortality, and the mean blood pressure of the population all increase with age; (2) hypertension is more common in young males than females, but in later years this relationship is reversed; (3) in the United States, hypertensive heart disease is more frequent among Negroes than among whites, and its mortality among both Negro males and females is higher than that among white males and females for every age group between 20 and 80; (4) the occurrence of hypertension (however defined) is higher among ethnic minority groups; (5) individuals in occupations requiring considerable physical exertion are more prone to develop hypertension than those in less strenuous occupations; (6) individuals in occupations involving exposure to specific metallic substances are more prone to hypertension; (7) individuals in occupations involving mental stress are more likely to develop hypertension.

Genetic or hereditary hypotheses: The evidence on the relationship of heredity to hypertension clearly indicates that there is a familial tendency in the occurrence of hypertension. There has been no final determination on whether the familial occurrence of hypertension is on a genetic or environmental basis, but the relationships are such that a specific genetic hypothesis can be formulated readily. This can be explored in a long-term study by comparing observations on different generations of the same families included in the study.

Constitutional hypotheses: Specific hypothesis formulation in this category is rendered difficult by the lack of conclusive evidence that body build as a whole or specific body characteristics are related to the development of hypertension. Such factors as body constitution, obesity, weight, and height have been examined, but more refined and precise measurements are needed if the relationships of these factors are to be explored in a long-term study. At present it appears necessary to take constitutional factors into account only to the extent of collecting such data as weight and height and perhaps a simple measure of body type. More refined measurements may be introduced after the study is under way, if there is suggestion from the simple measurements that the "constitutional" hypothesis is justified.

Nutritional hypotheses: Present opinion concerning the hypotheses in this category are expressed in Schroeder's[15] conclusion that "within the grounds of

present knowledge there is no evidence that human hypertension is caused by high protein diets, high fat diets, not enough vitamins or carbohydrates from the wrong source. . . ." Pending new leads, the testing of nutritional hypotheses is not a major item for long-term investigation. The design should provide for the collection of a minimal amount of nutritional data, including, perhaps, salt consumption, and should be flexible enough for additions, if indicated, during the conduct of the study.

Medical and physiologic hypotheses: These factors, considered at the Symposium,[14] have been reviewed in connection with the blood pressure measurement and hyperreactivity analyses.[2-4] Medical and laboratory examinations will constitute a part of the periodic observations of each subject in order to test specific hypotheses in this category and to determine the medical factors associated with the onset of hypertension.

OL.
4
956

It is recognized that the conditions of a long-term study might necessarily preclude the use of elaborate laboratory tests or clinical procedures. It is anticipated, however, that certain parts of the sample may be given appropriate tests from time to time as special projects within the long-term study. Furthermore, ethical considerations make it essential that when evidence of hypertension appears in an individual he must be given the benefit of appropriate examinations and treatment. Another individual without such evidence should be withdrawn from the sample and similarly examined as a control.

IV. The testing of hypotheses.—The unit of observation will be individuals in families, chosen by appropriate methods of population sampling.[16] Whatever observations are planned, the first requirement before an effective design can be formulated is the development, if possible, of a standard method of blood pressure measurement for study use. The procedures followed in an attempt to meet this requirement have been outlined. At this time a final decision has not been reached as to the conditions under which blood pressure readings will be made on each unit of the sample from time to time in order to assure a representative picture of the true range of blood pressure at each observation. Measurements will be made at intervals frequent enough to detect significant changes in peripheral vascular resistance which may be related to the hypotheses previously stated.

Other observations to be a part of the design include: a measurement of vascular hyperreactivity; some measure of cardiac and renal function; observations on genetic and certain constitutional characteristics; minimal data on nutritional habits; observations and measurement of psychologic and medical factors; data on occupation and other personal and social characteristics; and observations on the physical and economic environment. These observations will be made in part at the outset and also at appropriate intervals throughout the study period. The study design will adhere to the epidemiologic principle of identifying both the conditions under which hypertension *does* develop and those under which it does *not* develop.

The size and composition of the sample will be determined scientifically in order to meet the requirements for significant and reliable results. In order to eliminate a major weakness of so much past research on hypertension (i.e., use of biased samples) the basic population sample must be representative and

include people chosen from all walks and conditions of life. However, in order to assure significant comparisons of important variables, certain elements of the population now considered to be at greater risk will be included in a larger proportion in the sample than exists in the population (ethnic groups, occupational groups at risk, groups with family tendency to hypertension, etc.). Because of the long time involved in a study of this kind, the investigation should not be started until it can be assured that available funds will permit a sample of sufficient size to provide significant results.

Since cause and effect are a major concern of this study, adequate controls are essential. Because of the lack of a standard method of blood pressure measurement and the lack of acceptable criteria for defining hypertension, it is not possible at the outset to divide the sample into two groups: observations on nonhypertensives serving as controls for the observations on hypertensives. However, with similar observations on each unit of the sample, those not developing hypertension after a period of years can serve as controls for those who do, and the attending circumstances observed from time to time can be compared in the two groups in retrospect. It may be possible at the end of several years to match a control and study group according to age, sex, race, patterns of blood pressure variability, etc., and to observe them prospectively in the remaining years of the study for differences in outcome.

V. Conclusions and Practical Application.—This final step of an epidemiologic approach obviously cannot be fully realized until the end of the investigation and following an analysis of its results in terms of the underlying objectives, although interim reports may have considerable usefulness. The objective of an epidemiologic study, as an adjunct to laboratory and clinical research, is to determine the etiologic factors in disease development in order to suggest or develop means of prevention and control. By studying the conditions of disease occurrence, particularly in the phase of prepathogenesis, the etiologic factors in natural history may be elucidated and suggestions made as to how the natural history may be interrupted in man's favor. This is the practical application.

SUMMARY AND CONCLUSIONS

1. This paper reports the development, accomplishments, and future plans of a program of planning and research in the natural history of essential hypertension. Clinical and laboratory studies of hypertension have concentrated on the course of the disease in man but very little is known about the factors which initiate pathogenesis. It is the function of epidemiology to elucidate etiologic factors in the natural history of disease.

2. The development of the program is an illustration of the five steps of an epidemiologic method: (a) *the definition of the problem* was accomplished by a review of incidence, prevalence, and trend, and the objective stated at the outset was the establishment of a design for a long-term study of blood pressure in a population group in order to elucidate causative factors of essential hypertension; (b) *the appraisal of existing information* entailed the production of a bibliography on blood pressure of some 16,000 references and the organization of a symposium of experts to review present knowledge and indicate gaps in information which might be filled by a long-term epidemiologic field study; (c) ***hypothe-***

ses were formulated to determine the data which should be collected in the study and plans were developed to test the research procedures needed to collect the pertinent data; (d) *the actual test of the hypotheses* lies in the plans for the long-term study, which are briefly outlined; (e) the final step in the epidemiologic method, *conclusions and practical application*, must await the outcome of the studies, but, as with other epidemiologic studies, the expectations are for data which will aid in the prevention and control of essential hypertension.

3. The accomplishments of the program are reviewed herein and presented in detail in five papers which are to be published in sequence. These papers deal with blood pressure measurement, the phenomenon of vascular hyperreactivity, psychiatric and psychologic factors in hypertension, and racial factors in morbidity and mortality of hypertension.

4. The future plans of the program are described in a discussion of the elements of a long-term study and of the minimal requirements to be satisfied before a sound and comprehensive design can be formulated. Plans for meeting these requirements include further critical appraisal of existing data and two small-scale pilot field studies further to evaluate and test procedures for use in a long-term study.

VOL.
4
1956

REFERENCES

1. Clark, E. G.: Modern Concepts of Epidemiology, Editorial, J. CHRON. DIS. **2**:593, 1955
2. Glock, C. Y., Vought, R. L., Clark, E. G., and Schweitzer, M. D.: Studies in Hypertension. II. Variability of Daily Blood Pressure Measurements on the Same Individuals During a Three-Week Period, J. CHRON. DIS. (To be published.)
3. Clark, E. G., Schweitzer, M. D., Glock, C. Y., and Vought, R. L.: Studies in Hypertension. III. Analysis of Individual Blood Pressure Changes, J. CHRON. DIS. (To be published.)
4. Glock, C. Y., Vought, R. L., Schweitzer, M. D., Clark, E. G., and Katz, J.: Studies in Hypertension. IV. Comparison of Reactions of Three Tests for Hyperreactivity Among 204 Volunteers, J. CHRON. DIS. (To be published.)
5. Lennard, H. L., and Glock, C. Y.: Studies in Hypertension. V. Psychological Factors in Hypertension, An Interpretative Review, J. CHRON. DIS. (To be published.)
6. Glock, C. Y., and Lennard, H. L.: Studies in Hypertension. VI. Differences in the Distribution of Hypertension in Negroes and Whites. J. CHRON. DIS. (To be published.)
7. Leavell, H. R., and Clark, E. G.: Textbook of Preventive Medicine, New York, 1953, McGraw-Hill Book Company, Inc.
8. Perkins, W. H.: Cause and Prevention of Disease, Philadelphia, 1938, Lea & Febiger.
9. Evelyn, K. A.: The Natural Course of Essential Hypertension in Man, A Symposium on Essential Hypertension, An Epidemiologic Approach. Edited by the Division of Epidemiology and the Bureau of Applied Social Research, Columbia University, Boston, 1951, Wright and Potter, p. 63. (Reference 14.)
10. Epidemiology of Mental Disorder, New York, 1950, Milbank Memorial Fund.
11. Morsell, J. A.: The Problem of Hypertension: A Critical Review of the Literature Dealing With Its Extent, A Symposium on Essential Hypertension, An Epidemiological Approach. Edited by the Division of Epidemiology and the Bureau of Applied Social Research, Columbia University, Boston, 1951, Wright & Potter, p. 26. (Reference 14.)
12. A Bibliography of the World Literature on Blood Pressure, 1920-1950, (3 volumes). Edited by E. K. Koller and J. Katz. Boston, 1952, The Commonwealth of Massachusetts, Wright & Potter.
13. Clark, E. G., and Morsell, J. A.: An Epidemiologic Approach to the Study of High Blood Pressure, Am. J. Pub. Health **42**:542, 1952.
14. A Symposium on Essential Hypertension, An Epidemiologic Approach. Edited by the Division of Epidemiology and the Bureau of Applied Social Research, Columbia University, Boston, 1951, Wright & Potter.
15. Schroeder, H. A.: Hypertensive Diseases, Philadelphia, 1953, Lea & Febiger.
16. Merton, R. K., Morsell, J. A., and Clark, E. G.: A Proposal for Long Term Community Research in the Epidemiology of Hypertension, A Symposium on Essential Hypertension, An Epidemiologic Approach. Edited by the Division of Epidemiology and the Bureau of Applied Social Research, Columbia University, Boston, 1951, Wright & Potter, p. 327. (Reference 14.)

THE RISK OF BIASED SELECTION IN FORWARD-GOING SURVEYS WITH NONPROFESSIONAL INTERVIEWERS

DONALD MAINLAND, SC. D., M. B., CH. B., AND LEE HERRERA, B. S.

NEW YORK, N. Y.

*From the Department of Medical Statistics, New York University College of Medicine**

(Received for publication Feb. 26, 1956.)

VOL
4
195

WHEN a forward-going (longitudinal) survey is made in the search for an association between a disease and some characteristic, such as the smoking habit, large numbers of subjects must be enrolled if results are to be obtained within a short period. One method of meeting this difficulty, used by Hammond and Horn[1] in their study of smoking and death rates, was the employment of volunteer helpers, in this instance over 22,000 working members of local divisions of the American Cancer Society, who selected the subjects, handed out questionnaires, collected them, and reported on the subjects at yearly intervals.

TRAINING OF VOLUNTEERS

In the Hammond and Horn study, "the key volunteer leaders and many of the individual researchers were given verbal instructions by members of [the American Cancer Society statistical] office or by division office personnel as well as written instructions." The characteristics of a scientific investigation (objectivity, avoidance of biased sampling, the use of controls) are, however, not very easy to comprehend, for even a complete medical course does not always eradicate unscientific concepts regarding the requirements of valid evidence.

RISK OF BIASED SELECTION

It is conceivable, therefore, that some nonprofessional interviewers might believe that the best kind of subject in a habit-disease survey was one who showed both the habit and the disease. This does not imply a desire to distort the evidence, nor does it imply ability to diagnose a disease. It simply implies that some interviewers might feel that they would be doing a better job if they selected subjects who practiced the habit under investigation, and also had some symptoms attributable to the habit. In a smoking survey, the symptoms might be "smoker's cough" or "chest trouble." Perhaps even the knowledge that some person was under medical suspicion for a chest disorder might be considered a useful clue to a "good case."†

*550 First Avenue, New York 16, N. Y.

†During discussion of this problem, one of our colleagues, Miss M. I. Sutcliffe, who was at one time employed in commercial surveys, recalled how important it had seemed to the "researchers" to find evidence that would "please the boss."

It might be thought that in the Hammond and Horn smoking survey, this risk of bias could be discounted because "the volunteers ordinarily did not enroll men in the study who were seriously ill or dying at the time." This would account for the lower mortality in the subjects surveyed (smokers and nonsmokers) as compared with U. S. men in the same age groups[2] in national vital statistics reports; but the bias suggested here could operate equally well after the exclusion of the severely ill.

The lack of medical knowledge on the part of the interviewers cannot be held as a safeguard; for volunteer workers are often interested, enthusiastic, and persevering, and there are available books and medical friends to provide them with information regarding diseases and symptoms.

OL.
4
956

A SEARCH FOR POTENTIAL BIAS

It is difficult to see how the suggested bias could be detected during a survey; but it may be possible to form some estimate of the risk by asking potential interviewers their opinions as to what constitutes a "good subject" in such a survey. For this purpose we handed out a questionnaire to the New York University first-year medical class when it attended our first laboratory session in medical statistics at the beginning of the academic year 1955-1956. The admission regulations at this College are stringent, and it is safe to say that these students had been more rigorously selected, with reference to knowledge of science and intellectual acumen, than would be possible in securing volunteer helpers in a large survey.

It seemed desirable to issue the questionnaire before our teaching had influenced the students, and we believe that further time for thought and discussion among ourselves would have resulted in better questions; but the main results seem sufficiently clear to justify presentation.

QUESTIONNAIRE

Surveys on the Connection Between Habits and Disease

This is not a test for class grades; therefore there is no need to write your name on these sheets. However, please *remember* the number at the top right-hand corner for about a week, because we may wish to obtain some elucidation regarding an answer.

Large surveys are now being conducted in order to discover if there is a connection between diseases and habits like smoking and drinking. College graduates and others do "field work," i.e., asking questions or handing out printed questionnaires. One of the problems of organizers is to know how much, and what kind of, instruction must be given, and the following questions are an attempt to throw a little light on that problem.

It is important that each student answer *independently* of all other students. Therefore, please do not discuss this with your neighbors, or look at their answers.

Try to answer without asking the instructor anything; but if you have to ask, please raise your hand, and write your question, so that nobody will hear it.

Please give *your own opinion* in the answers, just as if you were about to start field work now. For simplicity of wording, drinking (liquor) will be used as an example. The information given to you by the organizer is as follows:

"You are to request 5 or 10 men (more if possible) to fill out questionnaires which ask about their present and past drinking habits. The men are to be between 50 and 69 years old. You will have to report on them annually for three or four years, stating whether they are alive or dead, and, if dead, where they died. Therefore you should ask the cooperation only of men with whom you expect to be in contact for the next several years.

"You should not select on the basis of drinking or not drinking, heavy drinking or light drinking, present drinking or past drinking."

In the following questions, a disease "attributable to drinking" means one that has been suggested, or believed, to result from drinking, although it can occur in people who have never drunk liquor.

"Valuable" means something (information or the selection of a person) that would be likely to be useful in answering the question at issue—the connection between disease and drinking.

A "nondrinker" means one who has never, or hardly ever, drunk liquor.

Circle in pencil the answer that expresses your opinion. Do not leave any question unanswered.

[In square brackets are the percentage frequencies for each answer. Total students = 129.]

Q. 1. Would it be valuable, if, before starting the field work, you learned something about the diseases (or their symptoms) attributable to drinking?
YES CERTAINLY [37.2] PROBABLY YES [17.1] CERTAINLY NOT [15.5]
PROBABLY NOT [30.2] NO OPINION [0.0]

Q. 2. Which, if either, would be more valuable?
A—A heavy drinker who is healthy.
B—A heavy drinker who is ill with a disease attributable to drinking.
CERTAINLY A [0.8] PROBABLY A [5.4] CERTAINLY B [1.6] PROBABLY B [7.8]
BOTH EQUALLY VALUABLE [80.6] BOTH OF LITTLE USE [1.6]
NO OPINION [2.3]

Q. 3. Which, if either, would be more valuable?
C—A heavy drinker who is healthy.
D—A heavy drinker who is ill with a disease *not* attributable to drinking.
CERTAINLY C [16.3] PROBABLY C [23.3] CERTAINLY D [0.8]
PROBABLY D [6.2] BOTH EQUALLY VALUABLE [45.0]
BOTH OF LITTLE USE [6.2] NO OPINION [2.3]

Q. 4. Which, if either, would be more valuable?
E—A heavy drinker who is ill with a disease attributable to drinking.
F—A nondrinker who is ill with a disease attributable to drinking.
CERTAINLY E [4.7] PROBABLY E [8.5] CERTAINLY F [7.0] PROBABLY F [6.2]
BOTH EQUALLY VALUABLE [69.0] BOTH OF LITTLE USE (3.9]
NO OPINION [0.8]

Q. 5. Which, if either, would be more valuable?
G—A heavy drinker who is ill with a disease *not* attributable to drinking.
H—A nondrinker who is ill with a disease *not* attributable to drinking.
CERTAINLY G [7.8] PROBABLY G [20.9] CERTAINLY H [0.8] PROBABLY H [2.3]
BOTH EQUALLY VALUABLE [33.3] BOTH OF LITTLE USE [33.3]
NO OPINION [1.6]

Q. 6. Have you, at any time, and anywhere, had a course in statistics?
YES [17.8] [NO 82.2]

INTERPRETATION OF RESULTS

Three students were absent, but all 129 students who were present answered all the questions, and they observed meticulously the request for independent work. It would perhaps be naïve to suppose that there was no guessing or haphazard answering; but at the beginning of their first year the students are much in earnest, especially regarding problems of disease. Although the two halves of the class attended the laboratory on successive days, they had little opportunity for full-class discussion during the interval. No question sheets were carried away, and the answers themselves created no suspicion of a different response between the two halves of the class.

The figures seem to indicate that a considerable number of students held opinions which, if allowed to influence the selection of subjects, would introduce bias into a survey. The risk would probably be greatest when the interviewer

VOL. 4 1956

met refusals to answer the questionnaire or delay in doing so. He would be tempted to give up the effort with subjects who, in his opinion, were not likely to contribute much of value to the survey, and he would be more inclined to persist with more "valuable" subjects.

The direction of the bias is a secondary matter—whether, for example, it would tend to create a preponderance of drinkers who are ill with a disease attributable to drinking (as suggested by the choice of B in Question 2), or whether some interviewers would introduce one kind of bias and others another kind. It is the evidence of *biased selection* of any kind that stamps the method as unreliable.

In order to estimate the proportion of students whose opinions, if applied in the selection of subjects, would entail a risk of bias, we interpreted, as revealing a definite risk, any questionnaire sheet that answered any or all of Questions 2, 4, and 5 by "Probably . . . " or "Certainly . . . " (followed by a letter).

If all of Questions 2, 3, 4, and 5 were answered by "Both equally valuable," we considered that there was no evidence of risk of bias.

In Question 3, a preference for C against D should probably also count as a risk of bias, because, although D was not "attributable to" drinking, in the sense defined in the questionnaire, an unbiased survey might reveal that it was associated with drinking. This question, we now realize, was not very well chosen, and speculation on what the answers might imply in certain situations would be profitless. The other questions provided sufficient evidence, and, therefore, to avoid exaggerating our estimate of the risk, the sheets that recorded preference for C, if answers to other questions had not shown bias, were classified as "Doubtful." (All sheets that showed preference for D were in the class of definite bias, because of answers to other questions.)

In any of the questions, an answer "Both of little use" would probably not entail risk in the particular choice under discussion, but it might well entail bias by excluding two classes of subject in favor of some other classes. Again, however, such an answer consigned the sheet to the "Doubtful" group unless it had been classified as biased for other reasons. The result of the classification was as follows:

	NUMBER OF STUDENTS	PERCENTAGE
Definite risk of bias	71	55.0
Doubtful	28	21.7
No evidence of bias	28	21.7
No opinion	2	1.6
	129	100.0

There was no association between these frequencies and the reported previous experience in statistics (Question 6).

The actual percentages are of some interest, but bias need be introduced by very few interviewers, each with only a few subjects, to produce in the analysis of survey results a "highly significant difference" between, for example, the mortality of drinkers and nondrinkers. Even if obtained from actual interviewers, such percentages could not be safely used to adjust the survey figures.

THE PROBLEM OF AVOIDING BIAS

The problem of how to instruct prospective interviewers to avoid bias is obviously difficult. If they were told to avoid all persons with illnesses, signs, or symptoms, this itself would introduce a bias which years of follow-up might fail to correct. We ourselves would relish neither the prospect of issuing instructions nor the task of making an unbiased selection from our acquaintances as subjects for such a survey. Indeed, the basic fault lies not in the interviewer, but in the nonautomatic selection of subjects. The hope that effects of biased selection will wear off[2] before losses from the sample are too great is insubstantial, especially in diseases whose symptoms last a long time before death.

We are not aware that any automatic method has been shown, in such surveys, to be a safe substitute for "probability" sampling, either simple or stratified.[3] It might be suggested, for instance, that after a certain hour on a certain day, predetermined by a supervisor, each interviewer select the first ten male acquaintances whom he met and who were in the required racial and age class. Besides introducing operational difficulties, this method would obviously tend to "cluster" sampling—the collection of subjects who were alike in many ways, such as socioeconomic level, type of employment, mode of life, habits (including the one under survey), and also exposure to risk of disease.

However desirable a piece of epidemiologic information may be, we may have to recognize that it is not merely difficult or expensive, but actually impossible to obtain. The reason may be the absence of a safe sampling method, the likelihood of refusals to answer questions, the prospect of losses in follow-up, or the impossibility of reliable classification of persons (e. g., by history, habits, or disease). Unless thorough and prolonged thought and detailed planning[4] give reasonable assurance of success, it is better to refrain from a survey than to produce equivocal results which not only waste effort and money but bring epidemiologic methods into disrepute.

SUMMARY

A questionnaire, given to a class of 129 first-year medical students, revealed that more than one-half of them held opinions which, if allowed to influence selection of subjects in a forward-going etiologic survey, would bias the results. This indicates a risk of bias in the selection by nonprofessional interviewers, but there appears to be no possibility of eliminating bias by training. The only safe method is automatic selection.

REFERENCES

1. Hammond, E. C., and Horn, D.: The Relationship Between Human Smoking Habits and Death Rates, J.A.M.A. **155**:1316, 1954.
2. Berkson, J.: The Statistical Study of Association Between Smoking and Lung Cancer, Proc. Staff Meet., Mayo Clin. **30**:319, 1955.
3. Committee on Sampling Techniques, American Public Health Association. On the Use of Sampling in the Field of Public Health, Am. J. Pub. Health **44**:719, 1954.
4. Hill, A. B.: General Principles of Field Surveys in Medical Research Council: The Application of Scientific Methods to Industrial and Service Medicine, London, 1951, H. M. Stationery Office, p. 7.

VOL. 4 1956

A COMPARISON OF SPECIFIC SYMPTOM DATA OBTAINED BY NONMEDICAL INTERVIEWERS AND BY PHYSICIANS

SIDNEY COBB, M.D., M.P.H., AND JOSEPH ROSENBAUM, B.A.

PITTSBURGH, PA.

From the Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh

(Received for publication May 10, 1956.)

THE extent of agreement between the nonmedical interviewer's reports of arthritis and rheumatism and the physician's diagnoses of these conditions has been reported.[1] This agreement was found to be poor and at the conclusion of the discussion it was stated, ". . . perhaps our objectives should be modified toward symptoms and disabilities in preference to specific diagnostic categories in order to permit preparation of estimates (i.e., estimates of prevalence rates) whose validity can be supported. . . ." It was assumed that much of the poor agreement observed between the physicians' diagnoses and answers obtained by the nonmedical interviewer was due to the vagueness of the terms "arthritis and rheumatism." In the present report an examination is made of the agreement between interviewers' and physicians' conclusions with regard to the history of such specific rheumatic manifestations as joint pain, joint swelling, and morning stiffness. These manifestations differ from the previously discussed question about arthritis and rheumatism in that much less judgment on the part of the interviewee is required. Furthermore, knowledge thereof is the personal information of the individual being interviewed. Competence in language, sincerity, and a reasonable memory are the only requirements for intelligent and accurate responses to questions about these symptoms.

METHODS

A household morbidity survey was conducted in July, 1952,[2] using a random sample of 3,000 households drawn from the Arsenal Health District of Pittsburgh.[3] This yielded medical and social information about 7,132 persons, 15 years of age and over. These persons were classified into three strata, for the projected Arthritis Study, on the basis of information regarding rheumatism and arthritis obtained from their respective household spokesmen by nonmedical interviewers:

> Stratum 1 consisted of individuals who were recorded as having arthritis or rheumatism.

Publication No. 5 of the Pittsburgh Arthritis Study. The research here reported has been supported by Grant No. A-308 from the National Institute of Arthritis and Metabolic Diseases.

Stratum 2 consisted of individuals who were recorded free of arthritis or rheumatism but were reported to have one of the following: painful swollen joints, cramps in arms or legs, pains in back, weak or painful feet, lumbago, gout, muscular rheumatism, sacroiliac strain, neuritis in arms or legs, and growing pains.

Stratum 3 was made up of the remainder of the 7,132 persons who were not recorded as suffering from rheumatism, arthritis, or related manifestations.

The plans for the Arthritis Study were completed in the spring of 1953, nearly a year after this household survey. A detailed questionnaire had been assembled and an arthritis clinic staffed by three physicians was established. Then a random sample of persons was selected for each sex separately and within each of the strata described above. This resulted in a total sample of 798 persons. Sampling rates of 60 per cent for males and 30 per cent for females were employed in strata 1 and 2. Both men and women were sampled at the identical rate of 7 per cent* in stratum 3. Each person thus sampled was visited in his home by a nonmedical interviewer equipped with the detailed arthritis questionnaire. The interviewer was instructed to voice the questions as printed on the interview form and to record all replies on the same form. Seven hundred and twelve persons, or 89 per cent of the sample, were successfully interviewed, while the remaining 11 per cent were divided into 4 and a fraction per cent who refused the interview and 6 and a fraction per cent unavailable for interview. Persons were unavailable for interview because they had moved from Allegheny County, could no longer be located by the staff, or had died since the household survey of 1952. The individuals who gave an interview were urged to have an examination by one of the three physicians in the arthritis clinic. Four hundred and seventy-eight persons visited the clinic for a history and a physical examination in addition to the interview. It is with these 478 people that we are here concerned. Obviously there was a lapse of time between the interview and the clinic visit. The median interval was two months with the lower quartile at one month and the upper quartile at six months. In order to compare the physicians' data with those obtained by the interviewers, only those cases were considered positive for the former when the onset of the symptom, as recorded in the physician's history, was either clearly antecedent to the date of the interview or in a period of time the major part of which preceded the interview. The reader should bear this in mind when inspecting the tables that follow, for the physicians' estimate of the prevalence of these symptoms prior to the time of interview is somewhat less than their estimate of prevalence at the time of examination, simply because the persons under study had aged in the interval. An examination of the agreement between physician and interviewer by interval between interview and examination did not reveal any significant improvement as the period grew shorter.

*The age group (15 to 34) in stratum 3 was actually sampled at the rate of 3½ per cent. In order to facilitate the analysis of data, these persons were given twice the weight of the remainder of stratum 3. There were 45 men and 46 women so treated and this increased the actual number sampled (707) to 798, and the actual number examined (429) to 478. Table I presents the breakdown of the weighted numbers, since these are the numbers necessary in all later analyses and tables. Readers who consult our earlier paper[1] will find the figures 707 and 429 used for this group of persons.

TABLE I. COMPARISON BETWEEN PHYSICIANS' AND INTERVIEWERS' ESTIMATES OF THE PROPORTION OF PERSONS HAVING CERTAIN RHEUMATIC MANIFESTATIONS BY SEX AND STRATUM AS OBTAINED BY PHYSICIANS AS OPPOSED TO NONMEDICAL INTERVIEWERS

SOURCE OF DATA	MALES				FEMALES				MALES AND FEMALES
	STRATUM				STRATUM				STRATUM
	1	2	3	ALL[1]	1	2	3	ALL[1]	ALL[1]
A. JOINT PAIN									
Physician's history	.95	.71	.49	.52	.94	.83	.48	.54	.53
Interview	.91	.57	.46	.48	.93	.61	.34	.41	.44
B. MORNING STIFFNESS									
Physician's history[2]	.41	.27	.16	.18	.58	.46	.17	.23	.20
Interview	.51	.33	.21	.24	.72	.39	.28	.33	.28
C. JOINT SWELLING									
Physician's history[3]	.41	.12	.12	.14	.56	.20	.16	.20	.17
Interview	.31	.20	.11	.12	.65	.34	.16	.22	.17
D. NUMBER OF PERSONS STUDIED									
Persons seen by both physician and interviewer	24	51	112	237	72	41	128	241	478

[1]The method used to combine relative frequencies of symptoms here, and in all other tables, is that of weighting the frequency for a particular stratum and then adding the weighted numbers.

$$\text{The weight used for a particular stratum} = \frac{\text{population of the stratum}}{\text{population in all strata being considered}}$$

[2]Four cases about which the interviewer was uncertain, and twenty-seven cases about which the physician was uncertain were classified as free of morning stiffness.

[3]Forty-two cases, for which the physician was uncertain about the history of joint swelling, were treated as "no's."

RESULTS

In Table I are found comparisons of prevalence estimates, based on data obtained by interviewers on the one hand and by physicians on the other, for the several rheumatic manifestations under study. These comparisons are made by sex and stratum so that the reader may see the extent of the variation from one stratum to another. First, let us consider section A of this table, entitled

TABLE II. THE ESTIMATION OF THE PROPORTION OF FALSE POSITIVES AND NEGATIVES WITH RESPECT TO A HISTORY OF JOINT PAIN FOR MALES STRATUM ONE

INTERVIEWER	PHYSICIAN		TOTAL
	YES	NO	
Yes	65	2	67
No	5	2	7
Total	70	4	74

Proportion of false negatives, $\theta = \frac{5}{70} = 0.07$

Proportion of false positives, $\phi = \frac{2}{4} = 0.50$

"Joint Pain." It is obvious that the physician consistently obtained somewhat higher estimates than the interviewer. This is, of course, in part due to the difference in approach. The interviewer asked the question "Have you ever had pain in any joint?" and recorded "yes," or "no," or "don't know." The "don't knows" occurred in only two cases and were included among the "no's."

TABLE III. THE PROPORTION OF CLASSIFICATION ERRORS FOR JOINT PAIN BY SEX AND STRATUM (THE NUMBERS IN PARENTHESES ARE THE DENOMINATORS FOR THE PROPORTIONS GIVEN)

CLASSIFICATION ERRORS	MALES				FEMALES				MALES AND FEMALES
	STRATUM				STRATUM				STRATUM
	1	2	3	ALL	1	2	3	ALL	ALL
Proportion of false negatives θ	.07 (70)	.33 (36)	.36 (55)	.34 (161)	.06 (68)	.32 (34)	.41 (61)	.37 (163)	.36 (324)
Proportion of false positives ϕ	.50 (4)	.33 (15)	.28 (57)	.29 (76)	.75 (4)	.29 (7)	.10 (67)	.17 (78)	.23 (154)

In the case of the physician's history, the "don't knows" numbered seven and were also included among the "no's." Pain due to injury was placed in the "no" category in both interview and physician's history. Although the physician and interviewer asked virtually the same question, the physician, while taking a history, had more opportunity to refresh the memory of his interviewee; and even if he received a "no" answer to his question initially, later on, in another context, he could discover that the person had had joint pain of a nontraumatic variety.

Another approach to these data is to compare the classification in a fourfold table such as Table II. It is convenient to describe the relationships in tables such as this in terms of the proportions of false negatives and false positives. The proportion of false negatives is defined as the proportion of persons answering "yes" to the physician, who answered "no" to the interviewer. Similarly, the proportion of false positives is defined as the proportion of physician "no's" given the interview classification of "yes." The values of these parameters for each sex and stratum are presented in Table III. Here we see again that there is variation from one stratum to another, which is quite understandable in view of the way the strata are defined. The over-all values of 36 per cent false negatives and 23 per cent false positives are disturbing, to say the least.

The next symptom is morning stiffness. The specific question asked by the interviewer was "Do you wake up with stiffness or aching in your joints or muscles?" with a possible response of "yes," "no," or "don't know." If the person responded "yes," he was asked, "Does your condition change as the day progresses?" with possible responses of "no," "better," or "worse." Those sampled persons who responded "yes" to the first part of this query and "better" to the second part were recorded as having morning stiffness according to the individual interview, and if their response was otherwise they were recorded as free of morning stiffness.

Returning to Table I and examining section B, "Morning Stiffness," we again find differences between estimates made by the physician and those made by the interviewer. This time, it is from the interview data that the higher estimate is made. This is, of course, in part due to the fact that the 27 cases, or just under 5 per cent of the total, classified as doubtful by the physician, were grouped with those free of morning stiffness. If these doubtful cases were considered positive, the over-all physician estimate would more nearly approach that of the interviewer, but would not exceed it. It would, however, seem inappropriate to classify the four interviewer doubtfuls with the negatives and the 27 physician doubtfuls with the positives.

The question asked about joint swelling was "Have you ever had swelling in any joints?" with answers of "yes," "no," or "don't know." The physician subsequently asked much the same questions of those persons appearing for examination but, of course, was permitted to probe much more carefully and to ask questions which tended to eliminate edema due to causes other than arthritis from being reported as joint swelling. In section C of Table I we see that the over-all estimate is the same for the physician's history and the interview. However, for some strata, such as females stratum 3, the estimates are identical, while in others the estimates vary rather markedly, sometimes showing a higher prevalence estimate for the interviewer, as in females stratum 2, and sometimes indicating a higher prevalence estimate for the physician, as in males stratum 1.

On seeing that there is good agreement between the interviewer and the physician on the prevalence of this manifestation, one might suppose that the classification errors would be small. The values for the proportion of false negatives and the proportion of false positives have been calculated as for joint pain, and the over-all values are presented along with the over-all prevalence estimates in Table IV. Here, it is seen that such a supposition is entirely incorrect. The significance of this will be discussed.

For comparison, data on the question "Have you ever had arthritis or rheumatism or other disease of that type?" has been included in Table IV. It is immediately seen that the order of magnitude of the classification errors is the same for all four questions.

TABLE IV. COMPARISON OF FOUR QUESTIONS RELATED TO RHEUMATIC DISEASE WITH RESPECT TO INTERVIEW VERSUS PHYSICIAN ESTIMATES OF PREVALENCE AND THE PROPORTION OF FALSE NEGATIVES AND FALSE POSITIVES

QUESTION	PROPORTION OF PERSONS		PROPORTION OF FALSE NEGATIVES θ	PROPORTION OF FALSE POSITIVES ϕ
	INTERVIEW ESTIMATE	PHYSICIAN ESTIMATE		
Joint pain	.44	.53	.36	.23
Morning stiffness	.28	.20	.37	.20
Joint swelling	.17	.17	.44	.09
Arthritis or rheumatism	.21	.22	.42	.13

DISCUSSION

It is evident that we must reject the hypothesis that much of the disagreement between interviewer and physician on "arthritis and rheumatism" was due to the vagueness of the terms "arthritis" and "rheumatism," for, superficially at least, there is nothing vague about joint swelling, which has proportions of false negatives and false positives which are almost identical with those for arthritis and rheumatism. At the moment we have no evidence as to whether the observed differences are principally due to differences in anamnestic skill between the physician and the interviewer or to classification differences or to variations in the responsiveness of the interviewee. Undoubtedly all of these factors are operating, but evaluation of their relative importance will have to await carefully planned studies of inter- and intra-observer variation.

We continue to feel uncomfortable about prevalence estimates that are based on data with classification differences as large as these. In order to understand more fully our concern in this matter it is important to see the relationship between the proportions of false positives and negatives, and the prevalence estimates made from the interview data and the physician data, respectively.

In the statement that follows, we make the simplifying assumptions that interviewer and physician see the same individuals and that sampling error is absent. In accordance with the work of Bross,[4] let us use the following symbols:

p = the true rate for a symptom, i.e., the rate determined by the physician.
$q = 1 - p$
p' = the corresponding rate determined by the interviewer.
θ = the proportion of false negatives.
ϕ = the proportion of false positives.

Then, p and p′ are related through θ and ϕ by the following equation:

$$p' = p - \theta p + \phi q$$

In words, this equation states that the proportion of persons judged to have a symptom by the lay interviewer is equal to the true proportion of such persons less the proportion falsely classified

as negative plus that falsely classified as positive. Inspection of the equation shows that the difference between the interview determined rate, p′, and the physician rate, p, is equal to $q\phi - \theta p$. In statistical language, this says that the bias introduced by substituting the interview data for that of a physician depends on 3 factors, the true rate, p, and the relative frequency of false positives, ϕ, and false negatives, θ. The form of this bias points up the fact that the two rates could agree fairly well, say in a sample, even though sizable frequencies of classification error are present. When this happens it is usually said that the false positives are compensating for the false negatives, or that $\theta p = \phi q$. This occurred in the sample analyzed in this paper. For example, consider the data for joint swelling in Table IV. The physician and interviewer estimates are identical and the proportion of false negatives, $\theta = .44$, and the proportion of false positives, $\phi = .09$. The bias in this situation is zero because $\theta p = \phi q$ or $.44 \times .17 = .09 \times .83$, the approximate equality being exact to two decimal places. Another fact of interest is that even small values of ϕ, i.e., a small proportion of false positives, can cause a large bias when one is measuring the rate of a relatively rare attribute in a population. For example, if we were dealing with a symptom whose true rate was $p = 0.02$, and assumed $\theta = 0.10$ and $\phi = 0.05$, then substituting in equation (1) we get

$$p' = 0.02 - 0.10 \times 0.02 + 0.05 \times 0.98$$

and find that $p' = 0.07$, which is three and one-half times as large as the true rate of 0.02.

OL.
4
956

Another source of concern about the reliability of prevalence estimates from these kinds of data is based on a related phenomenon. Suppose, for example, that in the example just cited there should occur, between two studies, a change in the proportion of false positives from 0.05 to 0.08. This would further increase the error in the prevalence estimate, giving $p' = 0.1$, which might lead the investigator to believe that the disease under investigation had increased in prevalence by almost 50 per cent when, in fact, all that had happened was a minor change in the response characteristics of the population under study. The danger of assuming that the false positives compensate for the false negatives must never be overlooked.

In the estimation of prevalence it is difficult to decide how or, in fact, whether to divide the effort between techniques with low validity and high response rate, such as the interview, and techniques of high validity and low response rate, such as the clinical survey. This is a complicated problem for which no ready solution seems available. It is, however, worthy of serious attention by statisticians. In the meantime, the epidemiologist will undoubtedly continue to seek field study methods that have both high validity and high response rate.

SUMMARY AND CONCLUSIONS

1. Prevalence estimates for history of joint pain, joint swelling, and morning stiffness referred to a common date vary according to whether the information is gathered by a lay interviewer or a physician.

2. The frequency of classification differences are as large for the three questions studied here as for the question on arthritis and rheumatism that was reported previously. It is not clear what proportion of these differences is inherent in any history-taking process and how much is due to basic differences between interviewers and physicians. It would appear likely, however, that the problem of making a judgment as to the presence of arthritis or rheumatism forms an unimportant part of the difficulty in making a satisfactory response to the questions asked.

3. The influence of false positives and false negatives, respectively, on prevalence estimates has been pointed out, and the danger of assuming that they will consistently compensate each other has been emphasized.

REFERENCES

1. Cobb, S., Thompson, D. J., Rosenbaum, J., Warren, J. E., and Merchant, W. R.: On the Measurement of Prevalence of Arthritis and Rheumatism From Interview Data, J. CHRON. DIS. **3**:134, 1956.
2. Ciocco, A.: Family Studies in Pittsburgh, Pennsylvania, in "Research in Public Health," New York, 1952, Milbank Memorial Fund.
3. Horvitz, D. G.: Sampling and Field Procedures of the Pittsburgh Morbidity Survey, Pub. Health Rep. **67**:1003, 1952.
4. Bross, I.: Misclassification in 2X2 Tables, Biometrics **10**:478, 1954.

VOL.
4
1956

THE USE OF INTERVIEW DATA FOR THE DETECTION OF ASSOCIATIONS IN FIELD STUDIES

THEODORE RUBIN, PH.D., JOSEPH ROSENBAUM, A.B., AND SIDNEY COBB, M.D., M.P.H.

PITTSBURGH, PA.

From the Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh

(Received for publication May 11, 1956.)

IN ALL studies involving the participation of the general public the problem of the bias introduced by those who do not respond is ever present. This problem has been particularly acute when clinical examinations have been involved. In the Pittsburgh Arthritis Study only 478 (60 per cent) of the 798 people who were requested to attend our clinic actually were examined. Other studies in the United States using clinical examinations have had response rates which vary between 54 and 69 per cent.[1-4] In such situations, what confidence can be placed in any conclusions concerning the associations between certain environmental factors and a disease when these decisions are based on the data obtained from the respondents? Can it be assumed that those who do not respond are "similar" to those who do? The only thing that can be done is to compare the relative frequency distributions of the respondents and nonrespondents on those variables, such as age, sex, race, etc., where the information is available for both groups. When such comparisons show no striking differences, there is "comfort" in concluding that there is no bias. However, this does not prove that such bias is absent. Therefore, methods of circumventing the nonresponse difficulty in future studies aimed at detecting associations deserve consideration. This report describes the circumstances under which interview techniques might be used.

The main reason for our interest in the use of interviews is that high response rates can be attained. The second advantage is that the nontechnical staff required for an interview study is more readily obtained than the staff of physicians, nurses, and technicians required for a study involving clinical examinations. Finally, there is the advantage of cost. Our experience would indicate that the cost of an interview with a specified individual is one-tenth the cost of giving this individual a complete medical examination.

Despite the advantages attributed to the interview technique, care must be exercised when a decision concerning the use of examinations or interviews for a given study is to be made. The investigator must realize that there may be

Publication No. 6 of the Pittsburgh Arthritis Study. The research here reported has been supported by Grant No. A-308 from the National Institute of Arthritis and Metabolic Diseases. Read at the meetings of the American Public Health Association, November, 1955.

individuals who would report on an interview that they are diseased but who would be classified as free of disease by a physician after a complete medical examination; these are referred to as false positives. Similarly, there may be individuals who would report on an interview that they are not diseased but who would be classified as diseased by a physician; these are referred to as false negatives. Our experience indicates that such misclassifications present a definite problem.[5] Only when the investigator can measure how this problem will affect his results can he make a rational choice between the use of medical examinations and interviews. It is our present purpose to discuss the measurement of the effects of misclassifications on the detection of associations and to indicate a promising method of dealing with the problem of "false positives" and "false negatives" so that such a rational choice may be made.

TABLE I. DEFINITION OF ASSOCIATION

	"TRUE" DISEASE STATUS (DETERMINED BY PHYSICIAN)		TOTAL
	DISEASED (RHEUMATOID ARTHRITIS)	NOT DISEASED (NO (RHEUMATOID ARTHRITIS)	
Factor Present (Z = 1) (Smoker)	U	V	U + V
Factor Absent (Z = 0) (Nonsmoker)	W	X	W + X

Proportion of Diseased in the "Factor Present" or "Z = 1" Group =

$$p_1 = \frac{U}{U + V}$$

Proportion of Diseased in the "Factor Absent" or "Z = 0" Group =

$$p_0 = \frac{W}{W + X}$$

There is association between the disease and the factor if $p_1 > p_0$.

In discussing the effects of misclassifications on the detection of associations there are two preliminary points that must be considered. First, the meaning of the term "association" must be defined. In making this definition the term "truly diseased" will be used to indicate that an individual's disease status is determined by a physician after a complete medical examination. However, such a diagnosis does not necessarily represent the actual state of an individual; that is, doctors make mistakes. Since the complete physical examination by a physician is the best source for a diagnosis available to us, the physician's report is accepted as "true." A factor, Z, is defined as being associated with a disease in a given population if the proportion, p_1, of truly diseased individuals among those in the population in which the factor, Z, is present is greater than the proportion, p_0, of truly diseased individuals among those in the population in which the factor, Z, is absent. Thus, as an example, we would say that rheumatoid arthritis is associated with smoking if the proportion of arthritics among smokers is greater than the proportion of arthritics among nonsmokers. This relationship is presented in Table I.

Second, our attention must be directed to the measurement of the effect of misclassification on the detection of associations. Generally speaking, one tries to determine whether or not association between a factor and a disease is present by using data obtained from a small segment of the population and drawing an inference regarding association in the population from the sample results. If the segment is a random sample of the population, this inference is made with the help of statistical tests of significance. In making conclusions based on such tests, certain errors may be committed. The risks of making these errors are measured in terms of probability statements.[6,7] The types of errors that may be made and the probability statements related to these errors are summarized in Table II. The first kind of error is committed when one erroneously concludes that there is association present (i.e., $p_1 > p_0$ when, in fact, $p_1 = p_0$). The risk of making this type of error is referred to as the significance level. The second kind of error is committed when one erroneously concludes that association is not present (i.e., $p_1 = p_0$ when, in fact, $p_1 > p_0$). The probability of *avoiding* this error, i.e., of correctly concluding that association is present, is known as the power of the test. It is important to note that these errors arise because one is dealing with a sample of the population. They must, therefore, be considered even when a 100 per cent response rate for the sample is attained and the "true" disease status, as determined by a physician on a complete medical examination, is used.

TABLE II. ERRORS THAT MAY BE MADE WHEN TESTING FOR ASSOCIATIONS

KIND OF ERROR	ERRONEOUS CONCLUSION	FACT	PROBABILITY STATEMENT
I	Association present $p_1 > p_0$	No association present $p_1 = p_0$	Significance level
II	No association present $p_1 = p_0$	Association present $p_1 > p_0$	Power

When using interview data to determine disease status, one would think that the risks of reaching erroneous conclusions would differ from the risks entailed when using "true" data (assuming a 100 per cent response rate in both situations). These differences in risks constitute our measure of the effect of misclassification.

From the foregoing discussion it might appear that these differences in risks depend on the false positives and false negatives that appear in the sample. This is not the case. Rather, these differences are based on the false positives and false negatives in the population from which the sample is drawn. More specifically these differences are dependent on:

(a) *the proportion of false negatives*, θ,* is the number truly diseased, who are classified as not diseased by the interview, divided by the total truly diseased.

(b) *the proportion of false positives*, ϕ,* is the number of persons truly free of disease, who are classified by the interview as diseased, divided by the total truly free of disease.

*It should be noted that θ equals 1—Specificity and ϕ equals 1—Sensitivity as these terms have been used by Kurlander, Hill, and Enterline.[8]

These proportions are illustrated in tabular form in Table III. In actuality the differences in risks of reaching erroneous conclusions between using "true" data and using interview data depends on these proportions in a slightly more complicated manner than indicated above. The reason for this complication is fairly simple. In detecting associations, two subdivisions of the population are being considered; one subdivision includes all those individuals in which the factor is absent (the Z = 0 group); the second subdivision includes all those in which the factor is present (the Z = 1 group). In each of these subdivisions or groups, there are false positives and false negatives. Tables similar to Table III can be constructed within each of these two groups. The proportion of false negatives in Z = 1 (smoker) group would be indicated by $\theta_1 = \frac{c_1}{a_1 + c_1}$ and the proportion of false positives in the Z = 0 (nonsmoker) group would be indicated by $\phi_0 = \frac{b_0}{b_0 + d_0}$, etc.

Now, given certain values of θ_1, θ_0, ϕ_1, and ϕ_0, the differences in risks can be computed. Practical considerations, however, place a limit on the useful range of values of these four proportions. For one, the proportions of false positives among the true negatives in the Z = 1 group (smokers) and in the Z = 0 group (nonsmokers) must be equal; similarly the proportions of false negatives among the true positives in these two groups must be equal. Symbolically, this limitation can be written $\theta_1 = \theta_0 = \theta$ and $\phi_1 = \phi_0 = \phi$. The significance of this assumption and the reasons for it will be discussed later. Under this assumption, it can be shown (see appendix) that the difference between:

(a) the proportion of individuals classified as diseased by means of their replies to certain questions on the interview in the Z = 1 group, symbolized by p_1'

and

(b) the proportion of individuals classified as diseased by means of replies on interviews in the Z = 0 group, symbolized by p_0'

is a constant, $1 - \theta - \phi$, times the difference in the proportions p_1 and p_0 (see Table I). This statement can be briefly written as an equation:

$$(1) \quad p_1' - p_0' = (1 - \theta - \phi)(p_1 - p_0).$$

If $1 - \theta - \phi$ is equal to zero, then $p_1' - p_0'$ is zero regardless of the relationship between p_1 and p_0. Therefore, the further restriction that $1 - \theta - \phi$ be greater than zero is placed on the proportions, θ and ϕ. Under the aforementioned assumptions the following two highly important statements may now be made on the basis of an examination of equation 1:

(a) *association between interview determined "disease" and the factor, i.e., $p_1' > p_0'$, means that there is association between the true disease and the factor, i.e., $p_1 > p_0$;*

(b) *when association between the interview determined "disease" and the factor is absent, i.e., $p_1' = p_0'$, association between the true disease and the factor is absent, i.e., $p_1 = p_0$.*

These statements are of the utmost importance because they permit us to draw an inference regarding true association when using interview data. The inference is made in the following manner:

(a) one accepts the conclusion that true association is absent whenever one decides that $p_1' = p_0'$;

(b) the conclusion that true association is present is accepted whenever one decides that $p_1' > p_0'$.

TABLE III. THE RELATIONSHIP BETWEEN THE TRUE AND INTERVIEW CLASSIFICATIONS OF DISEASE STATUS IN A POPULATION

INTERVIEW DISEASE CLASSIFICATION	TRUE DISEASE CLASSIFICATION		TOTAL
	DISEASED	NOT DISEASED	
Diseased	a	b	a + b
Not diseased	c	d	c + d
Total	a + c	b + d	a + b + c + d

The Proportion of False Negatives, i.e., the proportion of individuals classified as not diseased by interview data among those truly diseased, =

$$\theta = \frac{c}{a + c}$$

The Proportion of False Positives, i.e., the proportion of individuals classified as diseased by interview data among those truly not diseased, =

$$\phi = \frac{b}{b + d}$$

If these two rules of inference are followed, then, erroneous conclusions concerning true association are made *only* when erroneous conclusions concerning the relationship between p_1' and p_0' are reached. In particular, one erroneously concludes that true association is absent *only* when the incorrect decision that $p_1' = p'$ is made. Similarly, within the restrictions to be discussed later, one erroneously concludes that true association is present *only* when the incorrect decision that $p_1' > p_0'$ is made. To clear up any possible misunderstanding an example may be considered. Suppose there is interest in determining whether or not smoking is associated with rheumatoid arthritis. Furthermore, suppose that there really is no such association in the population, i.e., that the proportion, p_1, of arthritis among smokers is equal to the proportion, p_0, of arthritis among nonsmokers. Samples of smokers and of nonsmokers are selected and those in the samples are interviewed. Assume that the observed data lead to the decision that p_1' is greater than p_0'. Therefore, according to the stated rules of inference we accept the incorrect conclusion that true association is present. An examination of equation 1 will show that an error has been made in concluding that $p' > p_0'$ and simultaneously the incorrect conclusion that true association is present has been reached. Therefore, the risks in arriving at erroneous conclusions concerning true association when using interview data may be studied by considering the risks involved in arriving at erroneous conclusions concerning the relationship between p_1' and p_0'.

Let us discuss first the risks involved in incorrectly concluding that association is present. If examinations are used to determine the disease status of the individuals in the sample, then there is a certain risk due to sampling variation of making the incorrect decision that association is present. As has been stated before, this risk is called the significance level of the test. Suppose this level is set at 5 per cent. What is this level when interview data are used to determine the disease status of the individuals in the sample? As has been stated previously, one erroneously concludes that "true" association is present only when he erroneously concludes that the proportion of individuals who are classified as diseased from interview data, in the $Z = 1$ group, p_1', is greater than the proportion of individuals who are classified as diseased from interview data in the $Z = 0$ group, p_0'. However, this error is made only 5 per cent of the time. Again, sampling variation is responsible for this error. Thus, the significance level of the test for association between the "true" disease and the factor, using interview data to determine disease status, is 5 per cent. Thus, it is seen that the significance level of the test in the two situations is the same. A more mathematically sophisticated statement of this fact is included in the appendix. The credit for the original proof and for many of the ideas underlying this presentation belongs to Bross.[6]

The risk of erroneously concluding that association is absent remains to be considered. In this case there is a change in the risks. A simple example for illustrative purposes may be pro-

vided. Suppose p_1 = proportion of truly diseased individuals in the Z = 1 group (smokers) = 0.5, while p_0 = proportion of truly diseased individuals in the Z = 0 group (nonsmokers) = 0.1. The probability of coming to a correct conclusion that $p_1 > p_0$, which is called the power in our terminology, would then be quite high for the difference in proportions is large. However, suppose the interview classification was such that the proportion of false negatives among the true positives was 0.45, i.e., $\theta = 0.45$, and the proportion of false positives among the true negatives was 0.35,

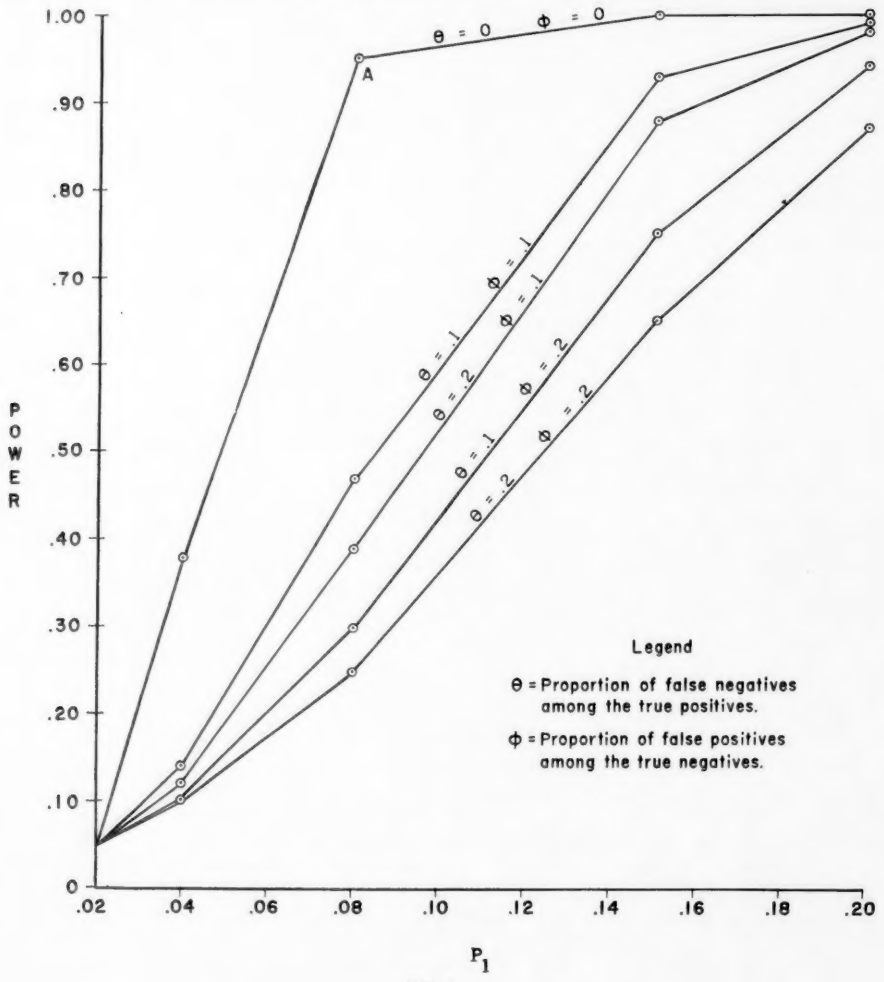

Fig. 1.

i.e., $\phi = 0.35$. Then by equation (1), $p_1' - p_0' = (p_1 - p_0)(1 - \theta - \phi)$, the difference in the proportions of individuals who would be classified as diseased via interview data in the two groups would be $0.4 \times 0.2 = 0.08$ (as opposed to a difference of 0.4 in the proportions of truly diseased individuals in the two groups). It should be evident that one is more likely to conclude that $p_1 = p_0$ using interview data, where there is a difference of 0.08, than when using true (examination) data, where there is a difference of 0.4 if the same sized sample is used in both situations. Therefore, the power (which is the chance or probability of reaching the correct conclusion that $p_1 > p_0$, i.e., that association is present) should be smaller when using interview data than when using true data if the sample size is the same in both situations. In Fig. 1 there is presented in graphical

form the power (the probability of correctly concluding that association is present) of the test as a function of θ, the proportion of false negatives among the true positives, and ϕ, the proportion of false positives among the true negatives, for a particular situation. In this case, the proportion of truly diseased individuals among those in the population in which the factor is absent, p_0, has been taken to be 0.02 (which roughly corresponds to our estimate of the prevalence of rheumatoid arthritis in the general population). The total sample size has been taken to be 500 where a sample of size 250 has been taken from each of the $Z = 0$ and $Z = 1$ segments of the population. The uppermost curve in Fig. 1 represents the situation when $\theta = 0$ and $\phi = 0$, i.e., the situation where there are no false positives and no false negatives. Hence, this uppermost curve is the power for the case where examinations are used to determine disease status. It also

TABLE IV. VALUES OF THE RATIO OF SAMPLE SIZE FOR INTERVIEW DATA TO SAMPLE SIZE FOR TRUE DATA REQUIRED TO GIVE THE SAME POWER FOR DIFFERENT VALUES OF p_0, p_1, θ, AND ϕ

ϕ PROPORTION OF FALSE POSITIVES	p_0*	p_1†	θ = PROPORTION OF FALSE NEGATIVES				
			0.05	.1	.2	.3	.5
0.05	.02	.04				5.39	10.35
		.08				4.12	7.71
		.15				3.32	5.99
	.05	.20				3.05	5.43
		.10				3.12	5.62
		.20				2.64	4.62
		.35				2.40	4.14
		.50				2.32	4.02
0.10	.02	.04		6.02	7.69	10.16	22.44
		.08		4.35	5.53	7.27	15.48
		.15		3.31	4.16	5.44	11.16
	.05	.20		2.94	3.69	4.80	9.73
		.10		3.11	3.92	5.08	10.55
		.20		2.47	3.09	3.98	7.98
		.35		2.11	2.63	3.39	6.65
		.50		1.95	2.44	3.15	6.15
0.20	.02	.04		12.62	16.48		
		.08		8.55	11.39		
		.15		6.06	8.05		
	.05	.20		5.17	6.86		
		.10		5.72	7.62		
		.20		4.19	5.55		
		.35		3.31	4.38		
		.50		2.90	3.86		
0.30	.02	.04	18.17	21.24			
		.08	12.27	14.37			
		.15	8.37	9.78			
	.05	.20	7.02	8.19			
		.10	7.94	9.31			
		.20	5.59	6.53			
		.35	4.22	4.97			
		.50	3.55	4.17			

*P_0 = Proportion of truly diseased among those free of the characteristic.
†P_1 = Proportion of truly diseased among those with the characteristic.

represents the improbable case where there is perfect agreement between the true and the interview disease status classifications. Even for small proportions of false positives and false negatives, e.g., $\theta = 0.1$, $\phi = 0.1$, one may see from Fig. 1 that there is a marked effect on the power over a wide range of p_1, the proportion of "truly" diseased among those in which the factor is present. Thus, by looking at Fig. 1, for $\theta = 0$, $\phi = 0$, the power when $p_1 = 0.08$ is about 0.95,

while for $p_1 = 0.08$, $\theta = 0.1$, $\phi = 0.1$ the power is about 0.45. Thus, for this value of p_1 the use of interview rather than examination data implies a twofold decrease in power even when the interviews present fairly good data. The procedure used to calculate these power functions and the theoretical basis for them may be found in the book by Eisenhardt, Hastay, and Wallis.[7]

Since we can calculate the power for a fixed sample size, we can also compute the sample size required for a given power. We have actually gone one step further and calculated the ratio of sample size for interview data to the sample size for true data to produce the same power for the interview as for true data. Selected values of this ratio are presented in Table IV. To illustrate the use of the table an example may be provided. Suppose we were using in an interview situation a rather good set of questions for which $\theta = 0.1$ and $\phi = 0.1$. Furthermore, suppose that $p_0 = 0.02$ and $p_1 = 0.08$. The value of the ratio in Table IV for this situation is 4.35. Therefore, if we use a sample of 435 and this set of questions on the interview instead of a sample of size 100 and examinations, the probability of reaching the proper conclusion that association is present will be unchanged.

Table IV may be used to illustrate another important point. In the range of prevalence that is relevant to disease studies, the value of the ratio herein tabulated is more strongly dependent on ϕ, the proportion of false positives, than on θ, the proportion of false negatives. Consider the situation where the proportion of truly diseased individuals among those in the population without the characteristic, p_0, is 0.02 and $p_1 = 0.04$. If the proportion of false negatives, θ, is 0.1 and the proportion of false positives, ϕ, is 0.1, the ratio has a value of approximately six; if $\theta = 0.3$ and ϕ remains 0.1, the value of this ratio increases to ten; but if θ remains 0.1 and ϕ is increased to 0.3, the value of this ratio is approximately twenty-one.

VOL 4 195

It has been demonstrated that within certain restrictions the significance level is not altered by using interview data in place of the more reliable examination data. Furthermore, it has been demonstrated that compensation can be made for the loss in power of the test, due to misclassification, by increasing the sample size; and a method of estimating the ratio of sample size needed for interview data to sample size for examination data has been presented. It is now logical to ask if questions or groups of questions such as may be used in interview surveys can be identified with ϕ and θ values low enough to give us practical values of this ratio. One convenient limit for practical values of the ratio is derived from financial considerations. In our experience it costs approximately ten times as much to do a thorough clinical work-up on an individual as to send an interviewer to his home. Therefore, it might be said that we are not anxious to work with this technique unless we can keep the value of this ratio below ten.

We are able to present some estimates of values of θ and ϕ because in the course of our field studies on arthritis we have had occasion to both interview and examine some 478 individuals. These individuals were classified by a physician as having classical or probable rheumatoid arthritis or as being free of the disease. They were also classified as having rheumatoid arthritis or as being free of the disease by means of their replies to certain questions on the interview. A fourfold table, such as Table III, was then set up for each question or group of questions considered and the estimates of θ and ϕ were computed. Some of these values are presented in Table V. Since ϕ is the more important of the two, as noted above, the questions and groups of questions have been arranged in order of decreasing value of ϕ. The bottom group of three questions is most suitable for the purpose at hand because it has the lowest value of ϕ ($\phi = 0.05$), and a fair value of θ (θ approximately 0.3).

Let us see how Table IV might be used to select a sample size for a future interview study of rheumatoid arthritis. First, it will be assumed that these estimates of θ and ϕ represent the true values. Second, some idea of the prevalence of rheumatoid arthritis in the $Z = 0$ group (the nonsmokers) must be

obtained. Our experience would indicate that for our definition of classical plus probable rheumatoid arthritis the prevalence in the general population of our study area lies in the neighborhood of 2.5 per cent of the persons who are over 14 years of age. Now, if a factor is associated with rheumatoid arthritis, one would expect that the proportion of diseased individuals among those without the characteristic, p_0, would be less than 0.025, perhaps 0.02. (Note that the proportion of diseased in the general population must be between p_0 and p_1.)

TABLE V. EVALUATION OF CERTAIN INDIVIDUAL INTERVIEW QUESTIONS AS SCREENS FOR CLASSICAL PLUS POSSIBLE RHEUMATOID ARTHRITIS

	ϕ PROPORTION OF FALSE POSITIVES	θ PROPORTION OF FALSE NEGATIVES
A positive or "don't know" response to:		
Have you ever had pain in any joints?	0.43	0.12
Do you wake up with stiffness or aching in your joints or muscles?	0.29	0.04
Have you ever had arthritis or rheumatism?	0.20	0.16
Have you ever had swelling in any joints?	0.16	0.27
Have you ever had any red joints that felt hot to touch?	0.07	0.90
Combinations of reports of:		
Morning stiffness and arthritis or rheumatism	0.12	0.18
Morning stiffness and joint swelling	0.09	0.31
Arthritis or rheumatism and joint swelling	0.09	0.31
Morning stiffness and arthritis or rheumatism and joint swelling	0.05	0.34

Decisions concerning the power and significance level must now be made. Suppose the significance level is set as 0.05 and the desire is nine times out of ten to detect a fourfold difference in the prevalences of rheumatoid arthritis between the group with the characteristic under study and the group free of this characteristic. For this particular instance, it can be computed that a sample of 250 persons should be drawn from each group when using the true classification. (This is the situation depicted by point A in Fig. 1.) Now, entering Table IV with the values $\phi = 0.05$ and $\theta = 0.3$, $p_0 = 0.02$ and $p_1 = 4 \times 0.02 = 0.08$, we come out with a value of the ratio of just over four. Our conclusion, then, is that in order to maintain the specified power of 0.9 we would need to interview a sample of 1,000 persons from the $Z = 1$ group and another 1,000 from the group free of the characteristic ($Z = 0$ group).

It should be noted that in doing this we have entered the table with the estimates of θ and ϕ obtained from our data. Since these are estimates and hence subject to sampling variation, it may be desirable to enter the table with slightly larger values of θ and ϕ in order to provide a margin of safety in the prediction that the power obtained will, in fact, be the required 0.9.

DISCUSSION

It was promised above that we would review the limitations of this method. They are six in number:

1. One of the assumptions used in the test was that $1 - \theta - \phi$ be greater than zero. This theoretical restriction requires that there be some degree of agreement between the interview and "true" classifications of disease status.

This is, therefore, a very mild restriction since interview data completely unrelated to the disease under study would not be used. However, practical considerations place a stronger restriction on the possible values of θ and ϕ than $1 - \theta - \phi > 0$. In particular, the values of θ and ϕ must be small enough to yield sensible values for sample size. As noted previously, ϕ has the greater influence on the required sample size when the disease under study has low prevalence. Therefore, it is particularly important that ϕ be small.

2. Generally speaking, the values of θ and ϕ, the proportions of false negatives and false positives, will not be known and hence must be estimated from the population under study. The decision as to whether or not the technique discussed here will be useful in a given study will depend strongly on the ease and accuracy with which θ and ϕ may be estimated. Clearly it is not sensible to do a large field study just to determine these quantities, for it should be remembered that these values primarily are sought as guides to assist our judgment in selecting a suitable sample size. There is, then, no point in determining ϕ and θ with great accuracy if only a vague notion of the general prevalence of the disease is available. Furthermore, there is no point in being very certain of these values if there is difficulty in deciding on the magnitude of the difference between p_1 and p_0 that should concern us. The requirement, then, is for estimates of these values that can be arrived at easily.

From definitions of θ and ϕ (see Table III) we see that θ is determined among persons who have the disease, while ϕ is determined among persons who do not. Since θ is less important than ϕ, in the determination of the necessary increase in sample size, it seems reasonable to determine θ in a population of persons preselected as having the disease, such as a group of persons seeking medical care in a hospital or clinic located in the area of the proposed study. This introduces a danger of bias, but the risk may be worth the price, for in studying diseases of low prevalence, large population samples are required in order to obtain a sufficient number of diseased persons on which an estimate of θ may be based.

On the other hand, it is probably better to estimate ϕ from a population sample. When most of the population is free of the disease under study, a comparatively small sample will be required. Each person in the sample should be seen by a physician in order to eliminate those with the disease. The physician's examination in most cases will be quite simple, for most of the persons examined will be obviously free of the disease, a few will have obvious involvement, and only a rare case will be on the borderline and so require a thorough examination.

3. There can be no relationship between the factor or characteristic under consideration and θ or ϕ. In other words, the possibility of a misclassification should not depend on whether or not the factor is present, i.e., $\theta_1 = \theta_0 = \theta$ and $\phi_1 = \phi_0 = \phi$. For example, let us assume that rural people are more apt to report arthritis when they have it than are urban people. This would result in a lower value of θ for the urban group, and hence there would be the possibility of a faulty conclusion that there is association between place of residence and arthritis, when, in fact, there is only association with a reporting characteristic. Another example of the introduction of association by differences in reporting characteristics in two divisions of the population (those with and those free of the

characteristic in question) is reported in Berkson's[9] noteworthy paper on the association between lung cancer and smoking. Our requirement that $\theta_1 = \theta_0 = \theta$ and $\phi_1 = \phi_0 = \phi$ prevents the introduction of this "false" type of association when no association is present in the population. If this assumption is not valid, then there is the strong possibility of reaching biased conclusions if the technique discussed in this paper is used. Since we originally considered this idea to avoid the possibility of a bias due to nonresponse, and since there is a possibility of another type of bias if the above assumption is not valid, it does not seem reasonable to use interview data when there is a reasonable suspicion that types of replies are correlated with the factor under study, without actually determining the values of θ and ϕ for the two groups. This is the most important of the limitations.

4. The ratios presented in Table IV, the power functions graphed in Fig. 1, and the theoretical presentation in the appendix apply only when one is able to draw samples from those who have the characteristic under study and from those who are free of this characteristic. On the face of it, this seems like a serious practical limitation. For one, the population is not very likely to be conveniently divided so that a sample may be taken from each of the two groups. Second, many factors may need consideration and it is impractical to draw two samples for every factor to be studied. However, there are two approaches to the problem. The first is a dodge to get around it. One can calculate the sample size needed as indicated above, estimate the frequencies of the characteristics in question, and then draw a large enough population sample to give, at least, the required number of individuals in each group. This, of course, becomes rather wasteful if the prevalence of one or more of the characteristics under study is small. Still, such an approach does seem reasonable, though the appropriate details of procedure and the theoretical consequences have not been fully explored.

The second approach is to extend the theory. One of us (J.R.) has been able to show that it is still true that the significance level is unchanged (under the assumptions $\theta_1 = \theta_0 = \theta$, $\phi_1 = \phi_0 = \phi$, and $1 - \theta - \phi > 0$) when the fourfold table (see Table I) has been obtained from a general population sample instead of samples taken from the segments of the population with and without the characteristic. The problem of the power functions in this new situation has not been considered as yet.

5. Pending further study it is probably safer to investigate only such variables as can be reasonably accurately determined, for we have made no attempt to study the problem of what happens when there is misclassification on both axes. We anticipate that this will be a difficult theoretical problem, but is worthy of serious attention.

6. It should be noted that special methods for estimating sample size are required when the sample is so small as to make it likely that there will be less than six cases therein contained. These methods are available.[10] The only point is that the ratios presented in Table IV are probably not valid for small samples.

With these limitations in mind the several advantages gained from the use of interviews may be emphasized. In the first place, interviewers can be hired

in any community and trained by the study staff, while physicians willing to do this sort of work are hard to find and usually also require a period of orientation to population studies. The second advantage is one of cost. As long as the ratio of the sample size for interview data to sample size for examination data is substantially below ten and the cost of estimating the proportions of false positives and false negatives, i.e., θ and ϕ, is reasonable, there is a financial saving in the use of interviewers.

The third and most important gain lies in the reduction of the nonresponse rate and, hence, a corresponding reduction in the possibility of a nonrespondent bias. As mentioned earlier, the response rates for clinical examinations in this country have been uniformly poor, while experience would indicate that completion rates of 90 per cent are attainable in interview surveys where the individual for interview is preselected.

Despite these advantages the investigator must exercise great care when he is making a choice between using interviews and using examinations. It is our hope that the notions presented here will aid him in making a rational choice by providing means of assessing the value of interview data for the study of associations. Furthermore, while this presentation has focused on the comparison of data from interviews and clinical examinations, the principles enunciated are applicable to many similar situations.

VOL 4 195

SUMMARY AND CONCLUSION

1. Studies involving clinical examinations of a sample of the general population have had response rates between 54 and 69 per cent. Such poor response rates brings the problem of a nonrespondent bias into the foreground. Methods of reducing the number, and proportion of nonrespondents must therefore be studied.

2. Our interest has centered on the use of interviews because interviews are inexpensive, are done with available personnel, and have high response rates.

3. The risks of arriving at erroneous conclusions concerning the association between a factor and a disease (a) when examination data are used and (b) interview data are used, have been considered.

4. Under certain assumptions, which have been stated and discussed, it has been shown (a) that the risk of erroneously concluding that association is present remains the same whether one uses interview or examination data; and (b) that the risk of erroneously concluding that association is absent is greater when using interview data than when using examination data.

5. Compensation for the greater risk of erroneously concluding that association is absent may be made by increasing the size of the sample used for interviews. A method for estimating the increase in sample size required is presented.

6. The restrictions on the use of interviews have been noted and areas requiring further theoretical work have been indicated.

7. It is concluded that there are circumstances under which associations can be detected by the use of interviewers with precision equal to or better than that attained by clinical examinations.

The assistance and guidance of Drs. Antonio Ciocco and Donovan Thompson of the Department of Biostatistics are gratefully acknowledged.

REFERENCES

1. Roberts, D.: Personal communication.
2. Tauber, J.: Prevalence of Heart Disease and Its Possible Relationships to Job Problems in an Adult Population. Thesis, University of Pittsburgh, 1955.
3. Trussell, R. E.: Personal communication.
4. Dawber, T. R., and Moore, F. E.: Longitudinal Study of Heart Disease in Framingham, Massachusetts: An Interim Report in "Research in Public Health," New York, 1952, Milbank Memorial Fund.
5. Cobb, S., Thompson, D. J., Rosenbaum, J., Warren, J. E., and Merchant, W. R.: On the Measurement of Prevalence of Arthritis and Rheumatism From Interview Data, J. CHRON. DIS. **3**:134, 1956.
6. Bross, I.: Misclassification in 2 x 2 Tables, Biometrics **10**:478, 1954.
7. Eisenhardt, C., Hastay, M. W., and Wallis, W. A.: Selected Techniques of Statistical Analysis, New York, 1947, McGraw-Hill Book Co., Inc.
8. Kurlander, A. B., Hill, E. H., and Enterline, P. E.: An Evaluation of Some Commonly Used Screening Tests for Heart Disease and Hypertension, J. CHRON. DIS. **2**:427, 1955.
9. Berkson, J.: The Statistical Study of Association Between Smoking and Lung Cancer, Proc. Staff Meet., Mayo Clin. **30**:319, 1955.
10. Cochran, W. G.: The X^2 Test of Goodness of Fit, Annals of Mathematical Statistics **23**:315, 1952.

APPENDIX

This appendix is a summary of theoretical considerations basic to the notions presented in the body of the paper. These considerations are in large measure derived from the presentation by Bross.[6]

Notation and Definitions.—

(1) "Truly diseased": an individual is "truly diseased" if he is classified as diseased by a physician after a complete medical examination;
(2) the Z = 1 group: those individuals in the population that have a given characteristic;
(3) the Z = 0 group: those individuals in the population that do not have this characteristic;
(4) p_1 : proportion of truly diseased individuals in the Z = 1 group;
(5) p_0 : proportion of truly diseased individuals in the Z = 0 group;
(6) p_1' : proportion of individuals classified as diseased from interview data in the Z = 1 group;
(7) p_0' : proportion of individuals classified as diseased using interview data in the Z = 0 group;
(8) A false positive: an individual who is truly free of disease but who is classified as diseased by means of interview data;
(9) A false negative: an individual who is truly diseased but who is classified as free of the disease by means of interview data;
(10) θ_1 : proportion of false negatives among true positives in the Z = 1 group;
(11) θ_0 : proportion of false negatives among true positives in the Z = 0 group;
(12) ϕ_1 : proportion of false positives among true negatives in the Z = 1 group;
(13) ϕ_0 : proportion of false positives among true negatives in the Z = 0 group;
(14) True association: $p_1 > p_0$;
(15) "Interview" association: $p_1' > p_0'$.

Assumptions.—

$$(1)\ \theta_1 = \theta_0 = \theta;\quad (2)\ \phi_1 = \phi_0 = \phi;\quad (3)\ 1 - \theta - \phi > 0.$$

Derivation of Equation.—

$$p_1' - p_0' = (1 - \theta - \phi)(p_1 - p_0)$$

Consider those in the Z = 1 group classified as diseased by means of the interview. These individuals are either truly positive or truly negative (the false positives). In the former group are those individuals classified as positive by both methods, i.e., those individuals truly positive that are not false negatives. Of the true positives the proportion, $1 - \theta_1$, is not falsely negative. Then, $p_1 (1 - \theta_1)$ is the proportion of persons in the Z = 1 group who are in this group. Now the proportion of false positives is $(1 - p_1) \phi_1$, where $1 - p_1$ is the proportion of true negatives in the Z = 1 group. Hence,

$$(1) \quad p_1' = p_1 (1 - \theta_1) + (1 - p_1) \phi_1.$$

Similarly,

$$(2) \quad p_0' = p_0 (1 - \theta_0) + (1 - p_0) \phi_0.$$

Under assumptions (1) and (2) $\theta_1 = \theta_0 = \theta$ and $\phi_1 = \phi_0 = \phi$.
Rewriting equations (1) and (2) and subtracting equation (2) from equation (1) yields:

$$(3) \quad p_1' - p_0' = (1 - \theta - \phi) (p_1 - p_0).$$

Now from equation (3), the following relationships should be quite obvious when assumption (3) $(1 - \theta - \phi > 0)$ is used:

(a) $p_1 - p_0 = 0$ implies $p_1' - p_0' = 0$, i.e., $p_1 = p_0$ implies $p_1' = p_0'$;
(b) $p_1' - p_0' = 0$ implies $p_1 - p_0 = 0$, i.e., $p_1' = p_0'$ implies $p_1 = p_0$;
(c) $p_1 > p_0$ implies $p_1' > p_0'$;
(d) $p_1' > p_0'$ implies $p_1 > p_0$.

Thus, the decision regarding $p_1 = p_0$ or $p_1 > p_0$ is determined by the decision between $p_1' = p_0'$ and $p_1' > p_0'$. Let us then look at the effect that using virtual data has on the significance level and the power of the test of the hypothesis that $p_1 = p_0$ against the alternative $p_1 > p_0$.

Significance Level.—

Suppose $p_1 = p_0$; then $p_1' = p_0'$. When one tests $p_1' = p_0'$ using the interview data, then the probability that one will reach the erroneous conclusion that $p_1' > p_0'$, is given by the significance level of the test. However, only when one reaches the erroneous conclusion that $p_1' > p_0'$ is the conclusion, also erroneous, that $p_1 > p_0$ is made. Therefore, the probability that one reaches the erroneous conclusion $p_1 > p_0$ is the same as the probability of reaching the erroneous conclusion that $p_1' > p_0'$. In other words, the significance level of the test is unaltered when interview data is used in place of true data.

Power.—

Suppose $p_1 = .08$ and $p_0 = .02$. Then $p_1 - p_0 = 0.06$. If virtual data for which $1 - \theta - \phi = 0.7$ were being used, then by equation (3), $p_1' - p_0' = 0.7 \times 0.06 = 0.042$. Thus, if the sample sizes were the same when using the different types of data, one would expect that he would come to the erroneous conclusion that $p_1 = p_0$ more often using virtual data than when using true data. Therefore, the power should be smaller when using interview data than when using true data. A glance at Fig. 1 will reveal that this supposition is borne out.

Required Increase in Sample Size.—

When there is a loss in power due to the classification error, it may be desirable to restore this power to the proposed level. This can be accomplished by enlarging the sample size. In Fig. 1 the power has been calculated for fixed sample size and various values of θ and ϕ. Similar computations can be made holding the power constant so as to give values of required sample size as a function of θ and ϕ. Or, better yet, ratios of sample size for virtual data to sample size for true data can be computed. This is what has been done for Table IV.

PROGNOSIS IN CHRONIC ULCERATIVE COLITIS

CHARLES A. FLOOD, M.D., MICHAEL J. LEPORE, M.D., ROBERT B. HIATT, M.D., AND AARON KARUSH, M.D.

NEW YORK, N. Y.

From the Departments of Medicine, Surgery, and Psychiatry, Columbia University, College of Physicians and Surgeons, New York, and The Presbyterian Hospital

(Received for publication March 3, 1956.)

VOL. 4 1956

THE long-term outlook of patients with nonspecific ulcerative colitis under conservative management is often unpredictable. The fact that few patients recover permanently has been documented repeatedly in large series of patients who have been resurveyed after various intervals of time.[1-6] Symptomatic remission may be brief or prolonged. An apparent tendency toward improvement after various types of treatment renders the evaluation of newly proposed methods of therapy difficult. The variability in the clinical course of this disease also makes it difficult to decide when surgical intervention should be considered.

A study of the clinical course of ulcerative colitis was initiated by us seven years ago for the purpose of studying the prognosis in this disease in greater detail. It was hoped that such a study would provide a graphic representation of the continuous course of the disease which would be useful in the evaluation of methods of therapy, whether medical, psychiatric, or surgical.

The present report summarizes experience with a series of patients with ulcerative colitis involving the colon and rectum, who have been observed under conservative management during the seven-year period from 1948 to 1955. If the patients had been seen prior to that time, the previous course is also included in this analysis. The outlook of the average patient is evaluated. The relation between various clinical findings and prognosis is analyzed.

The results of the present study suggest that a number of clinical findings bear a significant relation to prognosis. Significant variations in long-term outlook appear to be related to such factors as the extent of anatomic involvement, the rate of symptomatic response to treatment, and the ability of the patient to manifest a complete remission in the inflammatory process in so far as it can be judged by sigmoidoscopic examination.

SERIES OF CASES

The present study is limited to patients with ulcerative colitis which involved both the rectum and the colon. All such patients are included who remained under observation for a minimum period of three months during the seven-year period from 1948 to 1955. Patients with ulcerative disease which appeared to be

limited to the rectum and rectosigmoid and patients with segmental colitis without rectal involvement are excluded from present consideration.

One hundred and forty-eight patients fulfilled these criteria. The series includes 69 males and 79 females. The age distribution of the patients at the time of the initial visit is shown in Table I.

TABLE I. AGE DISTRIBUTION OF PATIENTS IN STUDY

AGE IN YEARS	NUMBER OF PATIENTS
0– 9	8
10–19	27
20–29	41
30–39	32
40–49	22
50–59	11
60–69	2
70–79	3
80–89	2
Total	148

The duration of the disease at the time of the initial visit was relatively short in approximately half of the patients, being less than a year in 48 patients and less than two years in 28 other individuals. The history was dated back more than two years to over ten years in the remaining patients.

Two-thirds of the patients were considered ill enough to require hospital admission when they were first seen. The other one-third had only mild or moderate symptoms at the time of the initial visit and were placed on ambulatory treatment. Only two individuals were symptom-free at the time of their first visit.

The program of treatment corresponded to the type which is in general use at the present time. The chief therapeutic measures which were employed included a liberal low-residue diet, supplementary vitamins, mild sedatives, anticholinergic drugs, sulfa drugs, constipating substances such as aluminum hydroxide, and parenteral fluids. More than one-half of the patients received antiamoebic therapy. Antibiotics were administered to a large number of patients in the usual dosage. Steroid therapy or corticotropin was used in only a small proportion of the cases and only in selected patients whose condition was serious. The details of the latter group will be summarized later. Opiates and other narcotics were rarely employed except during hospitalization. The majority of the patients underwent psychiatric interviews and many individuals received formal psychotherapy for varying periods of time.

After discharge from the hospital the patients were usually seen at intervals which ranged from once every two weeks to once every three months, depending on the severity of symptoms. Clinic visits were not contingent upon the presence of symptoms.

ESTIMATION OF ACTIVITY OF DISEASE

For purposes of simplification, the symptomatic activity of the disease has been graphed for each patient chiefly in terms of the severity of the bowel symptoms. This provides a useful approximation of the clinical course. In general,

the amount of diarrhea reflects the severity of the general clinical picture, with the exception of a few complications such as arthritis and skin infection. Obviously several other symptoms might have been selected as indices of activity of chronic ulcerative colitis, such as weight loss and other constitutional effects of the disease. Since the inflammation of the colon is the outstanding aspect of the disease, it seems justifiable to attempt to estimate the severity in terms of the local symptoms and signs.

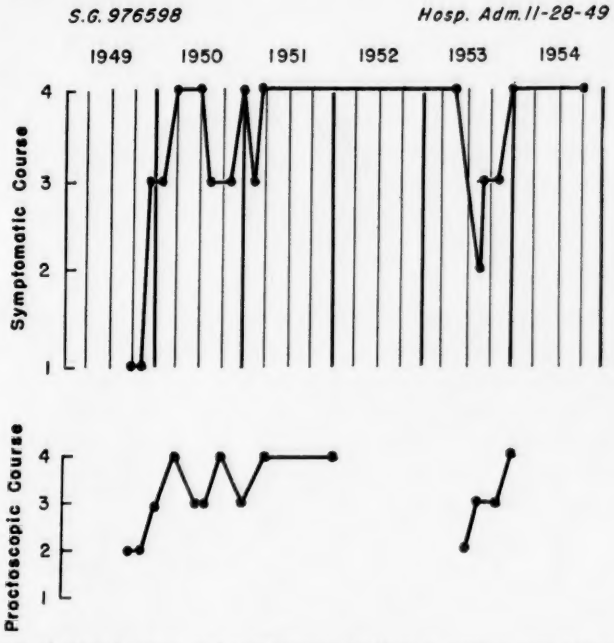

Fig. 1.—Graph of clinical course and proctoscopic findings in patient with ulcerative colitis first seen in 1949. Severity of symptoms is indicated above as follows: *1* = severe, *2* = moderate, *3* = mild, *4* = asymptomatic. A rough parallelism was noted between the proctoscopic changes and the symptomatic course.

The symptomatic course over a period of time has been graphed in the following manner (Fig. 1): *1* = incapacitating symptoms including diarrhea with stools averaging over ten per day; *2* = moderate symptoms—diarrhea of 6 to 10 stools daily; *3* = mild diarrhea with up to 5 stools daily or the presence of any rectal bleeding; and *4* = formed stools without blood. Fig. 1 illustrates graphically the course of a patient over a period of years. It has been possible to combine the results, expressed in this manner, in a series of patients for the purpose of studying the effect of various factors on the course of the disease.

Proctoscopic examination offers the most useful method of determining the presence of active inflammation in the bowel in ulcerative colitis.[2] Efforts to grade the severity of the inflammatory changes in the rectum yield variable results with different observers. We have, therefore, been satisfied to grade the proctitis as active = *2;* minimal or equivocal = *3;* and inactive = *4.* Initially, an attempt was made to grade the severity of the inflammatory process in greater detail but this proved to be impractical.

IMMEDIATE RESULTS

The results have been divided into the immediate and the long-term. The clinical course during the first three months of observation was arbitrarily selected for the estimation of the immediate results. The optimal symptomatic status achieved by the patients within the first three months of observation was termed the immediate result irrespective of whether improvement was transitory or sustained. The immediate results for the entire group of 148 patients are shown in Table II.

TABLE II. INCIDENCE OF EARLY SYMPTOMATIC REMISSION

NUMBER OF PATIENTS	4 (WELL)	3 (MILD DIARRHEA)	2 (MODERATE SYMPTOMS)	1 (SEVERE SYMPTOMS)
148	61 (41%)	35 (24%)	28 (19%)	24 (16%)

During the first three months of treatment 96 patients (65 per cent) had relatively satisfactory symptomatic results in that the symptoms either cleared up completely (*4*) or became minimal (*3*) in severity. In general, approximately two-thirds of our patients experienced early symptomatic improvement. Of the 24 patients who are classified symptomatically as *1*, that is, no remission whatever, 15 individuals were subjected to early operation and three other patients died under conservative management.

It should be emphasized that the improvement registered by these patients was often, and in fact usually, transitory. More than one-half of the patients again experienced some increase in symptoms within a short period after the initial improvement. For purposes of appraising the immediate results, an exacerbation occurring immediately after a short period of improvement has been disregarded.

LONG-TERM RESULTS

The long-term results were studied in those patients who remained under medical management for a period of time which exceeded six months. There were 113 such individuals in this series. Seventeen additional patients were lost to follow-up in less than six months and are therefore not included in the tabulation of long-term results. The average period of consecutive follow-up observation was 2.9 years. Only seven patients remained entirely free from symptoms throughout the period of follow-up observation. These seven patients have been followed for periods varying from one to three and one-half years.

An analysis of the results for the 113 patients in terms of the years in which symptoms were present or absent is set forth in Table III.

The symptomatic course was classified as satisfactory for approximately half of the years of follow-up observation. Forty-eight per cent of the years were classified either as years of complete freedom from symptoms (symptomatic *4*), or years in which the symptoms were minimal (symptomatic *3*).

VOL 4 195

Six deaths occurred during the follow-up period under conservative management. One death was due to carcinoma of the rectum. Two deaths were attributable chiefly to associated chronic liver disease. In two patients death seemed clearly attributable to other unrelated diseases, and the cause of death in one additional patient was undetermined.

TABLE III. LONG-TERM RESULTS BY YEARS

NUMBER OF PATIENTS	SATISFACTORY YEARS (SYMPTOMS ABSENT OR MINIMAL)		UNSATISFACTORY YEARS (MODERATE OR SEVERE SYMPTOMS)	
	4	3	2	1
113	83 (26%)	72 (22%)	95 (30%)	73 (22%)

Average follow-up—2.9 years.

Sixteen other patients were subjected to surgical intervention because of an unsatisfactory clinical course. Three of these patients died during the early postoperative period.

In summary, the immediate and the long-term results of the entire group of 148 patients were as follows: thirty-one individuals (20 per cent) ultimately came to operation and nine patients (6.1 per cent) died under conservative management. Early symptomatic improvement took place in approximately two-thirds of the patients. Long-term results under medical management were relatively satisfactory (symptoms remaining absent or minimal) during approximately one of every two years of subsequent observation.

RATE OF RESPONSE TO THERAPY

One of the most serious problems in management is the acutely ill patient who fails to improve during a period of several weeks of intensive medical treatment. In the present series, many such patients came to operation. For the purpose of studying this problem we have selected those patients who ran a daily febrile course for more than fourteen days, with oral or rectal temperatures which exceeded 103° F. on one or more occasions. There were twenty patients with records of a febrile course of this severity (Table IV). All except three patients received one or more sulfa drugs or antibiotics in the usual dosage during the febrile course. Most patients received several such drugs, including penicillin, Aureomycin, or Chloromycetin. No obvious effect of such therapy on the fever was apparent, although the treatment may have been beneficial in limiting the severity of the febrile response. Cortisone and corticotropin were not administered to any of these patients during the first two weeks of hospitalization. However, these substances were given to two patients during the third week of fever and were followed by prompt defervescence and symptomatic improvement.

TABLE IV. EARLY AND LATE RESULTS IN 20 PATIENTS WHO HAD FEVER FOR OVER 2 WEEKS, REACHING 103° F. ON ONE OR MORE OCCASIONS*

PATIENT	APPROXIMATE DURATION OF FEVER	HIGHEST TEMPERATURE (°F.)	EARLY RESULT (FIRST 3 MONTHS)	SUBSEQUENT COURSE	PRESENT STATUS
1	15 days	104	Colectomy after 6 weeks		
2	4 weeks	104	Colectomy after 4 weeks		
3	16 days	103	Colectomy after 2½ months		
4	10 weeks	103.6	Ileostomy at 2½ months		
5	3½ weeks	103.2	Ileostomy at 11 weeks		
6	4½ weeks	104.8	Cecostomy after 4 weeks		
7	3½ weeks	104.6	Vagotomy after 9 weeks		
8	8 weeks	105	Died after 8 weeks		
9	7½ weeks	104	Died after 7½ weeks		
10	4½ weeks	106.4	Died after 4½ weeks (also had lupus erythematosus)		
11	15 days	103	Partial remission (3)	Many exacerbations in next 5 years	Colectomy
12	2½ weeks	104	Partial remission (3)	Four severe recurrences in 5 years	Colectomy
13	2½ weeks	104.6	Complete remission (4)	Severe recurrence after 2 years	Colectomy
14	2 months	103.6	Partial remission (3)	Stormy course for 1½ years	Colectomy
15	7 weeks	104.4	Partial remission (3)	Severe recurrence after 5 months	Vagotomy
16	3 months	103	Unsatisfactory (2)	Many subsequent admissions	Colectomy
17	2 weeks	104.4	Complete remission (4)	Full remission for 2 years (4)	Asymptomatic
18	2½ weeks	107.4	Partial remission (3)	Satisfactory for 4½ years (4 or 3)	Asymptomatic
19	5 weeks	103	Complete remission (4)	Satisfactory for 7 years (4 or 3)	Asymptomatic
20	3 weeks	103	Complete remission (4)	One severe recurrence in 3 years	Asymptomatic

*Only 4 patients have had a comparatively satisfactory course under conservative management.

VOL. 4 1956

In general, the results in this group of twenty patients with high and prolonged fever were poor. Seven individuals were subjected to surgical intervention within three months. Three patients died under conservative treatment. Six of the remaining patients improved under medical management but later suffered repeated severe exacerbations and were eventually operated upon. Only four individuals now remain under medical therapy and have had a benign follow-up course. The extent of anatomic involvement demonstrable by x-ray examination in these twenty patients was quite variable. Thirteen individuals had only partial involvement of the colon which was demonstrable radiologically at the time of the acute febrile episode. Three of these individuals were found to have universal disease of the colon at a later date. Six patients had universal disease initially and one patient was not studied radiologically.

The less seriously ill patient who fails to go into at least a partial symptomatic remission under medical management also presents a difficult problem in management. A study was made of the long-term follow-up course of patients whose response was classified as continuously unsatisfactory during the first three months of treatment. Individuals who continued to have more than five stools daily (symptomatic *1* or *2*) during the entire first three months, fell into this group. There were 28 such patients who remained under observation for more than six months. Thirteen patients presented evidence of disease which involved the entire colon and fifteen patients had partial colonic involvement. The results are shown in Table V.

TABLE V. LONG-TERM RESULTS IN PATIENTS WITH DELAY IN IMPROVEMENT

NUMBER OF PATIENTS	SATISFACTORY YEARS		UNSATISFACTORY YEARS	
	4	3	2	1
28	3	7	29	27

Average follow-up—2.4 years.

In this small number of patients the proportion of unsatisfactory to satisfactory years was high. Approximately four out of five years were classified as unsatisfactory as contrasted to an average of one out of two unsatisfactory years for the entire series. However only three patients were ultimately subjected to surgery. One patient died under conservative treatment. Many of these patients disappeared from follow-up observation after a period, possibly because of discouraging results.

SIGNIFICANCE OF PROCTOSCOPIC REMISSION

The changes in the rectal mucosa which are seen on proctoscopic examination in nonspecific ulcerative colitis are extremely varied.[2] The commonest alterations include hyperemia, edema, oozing of blood from the mucosal surface, and patches of exudate. A minority of patients present ulcerations which are large enough to stand out discretely and which occasionally may reach a centimeter or

more in diameter. Diffuse granularity of the rectal surface is a common finding. The granular surface may tend to bleed freely although there is usually little or no gross surface exudate visible when the picture is of the granular type. The picture of exudative inflammation is seen commonly in the sicker patients. As symptomatic improvement occurs, the granular type of appearance may supervene. We have regarded all of the above changes as evidence of active disease.

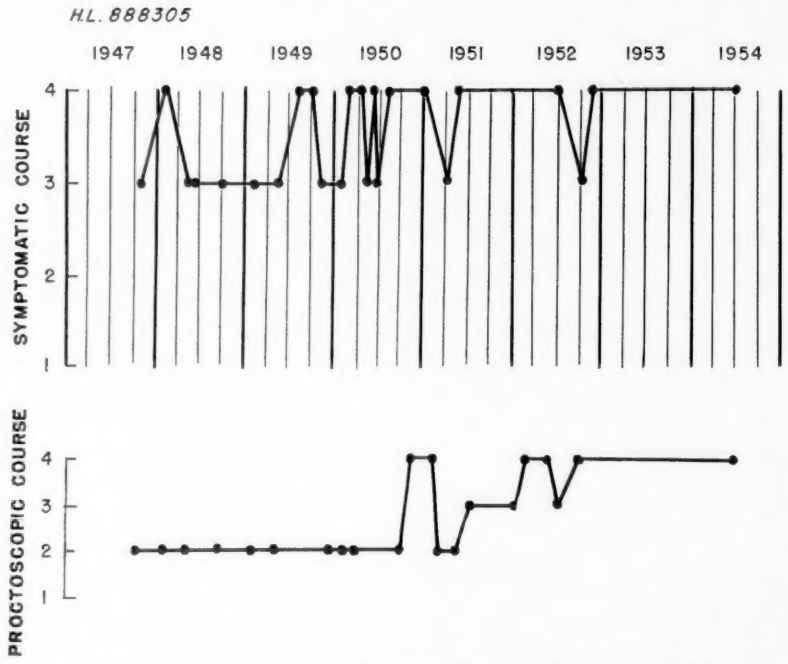

Fig. 2.—Graph of serial proctoscopic findings in patient with intermittent mild symptoms of ulcerative colitis. Frequent proctoscopic examinations revealed evidence of active inflammation throughout the first three years although the symptoms came and went. Finally, proctoscopic evidence of remission appeared (*2* = active inflammation, *3* = equivocal findings, *4* = no active disease).

Whenever the proctoscopic examination ceased to reveal changes which would substantiate a diagnosis of ulcerative colitis, it was considered that proctoscopic remission had occurred. In some cases, the rectum became entirely normal in appearance. Absence of active disease was also considered to be consistent with a proctoscopic examination which only revealed residuals of previous disease, such as a loss of the visible vascular pattern or residual mucosal pitting. A tubular inelastic appearance of the rectal lumen was also a not uncommon residual finding. The continuing presence of pseudopolyps in the rectum without active inflammatory changes was likewise not classified as evidence of active disease.

Not uncommonly the rectal mucosa approaches normal appearance but it is still possible to detect a few patchy areas of mucosal granularity. Such minimal changes are often found just prior to a complete remission of the inflammatory process in the rectum. This type of change was considered to be evidence of continuing active mild inflammation and was classified as *3* for purposes of graphing the course of the disease (Fig. 2).

Complete proctoscopic remission, as defined above, was observed in 38 patients in the present study. Serial proctoscopic examinations of some patients were carried out with sufficient frequency to permit an approximate determination of the length of time which elapsed from the beginning of medical supervision until such objective evidence of remission was apparent. Complete remission was observed within periods of one to three months in only five patients. The available data suggested that in most instances the subsidence of inflammation in the rectum took place more slowly and often required over six months. A number of patients in the series, who were proctoscoped frequently, presented evidence of a continuously active process for over two years before going into proctoscopic remission. It was noteworthy that the appearance of proctoscopic remission lagged many months or even years behind the subsidence of symptoms in most cases in which such remission occurred.

OL.
4
956

Proctoscopic remission was observed, at times, both in patients with total involvement of the colon as well as in patients with partial colonic involvement, as seen on radiologic examination. Ten of our patients who manifested proctoscopic remission had total colonic involvement while the other 28 patients had only partial colonic involvement.

The long-term symptomatic course subsequent to complete proctoscopic remission was strikingly better than the average curve of performance in this disease. Thirty-three patients were followed for an average period of 2.6 years after complete proctoscopic remission. The results appear in Table VI.

TABLE VI. SYMPTOMATIC COURSE FOLLOWING PROCTOSCOPIC REMISSION

NUMBER OF PATIENTS	SATISFACTORY YEARS		UNSATISFACTORY YEARS	
	4	3	2	1
38	39	27	13	6

Average follow-up—2.6 years.

Seventy-eight per cent of the follow-up years were relatively satisfactory in terms of symptoms.

EXTENT OF ANATOMIC INVOLVEMENT

Several anatomic factors are considered to have an unfavorable significance in relation to the prognosis in ulcerative colitis. These include extensive colonic involvement, involvement of the ileum, the presence of pseudopolypi, and the development of strictures. The results in the patients with these characteristics have been analyzed in the current study.

1. *Total Involvement of the Colon.*—Sixty-two patients presented evidence of inflammatory disease of the entire colon. The immediate results in these patients were less satisfactory than the average for the entire series. During the first three months of observation, only 50 per cent of these patients became, at least temporarily, symptom-free or improved to a point at which the symptoms were

classified as mild. During this period two patients died of the disease and nine other individuals were subjected to operation because of an unsatisfactory early clinical course.

Long-term observations were available for 44 patients with total colonic involvement who were followed for at least six months. The average period of observation was 2.5 years. The results were less satisfactory than in the full series as seen in Table VII.

TABLE VII. LONG-TERM RESULTS IN TOTAL COLONIC INVOLVEMENT

NUMBER OF PATIENTS	SATISFACTORY YEARS		UNSATISFACTORY YEARS	
	4	3	2	1
44	22 (20%)	15 (14%)	35 (32%)	38 (34%)

Average follow-up—2.5 years.

Only about one-third of the follow-up years were classified as satisfactory in the patients with total colonic involvement. Rehospitalization was frequently necessary in this group of individuals. Nine patients in this group were subjected to surgical treatment. Five others died, death being due to a complicating carcinoma of the rectum in one instance and, in part, to associated hepatic cirrhosis in two patients. One patient died of an incidental brain tumor, and the cause of death was unknown in one instance.

The more unsatisfactory clinical course reflects the increased seriousness of ulcerative colitis which involves the entire large bowel. It should be noted, nevertheless, that many patients with universal involvement may get along very well under conservative management for long periods of time.

2. *Ileal Involvement.*—Sixteen patients with total involvement of the colon were considered on radiologic examination to have ileal disease as well. In all but a few instances the visible changes were limited to the last few inches of the terminal ileum. Ten of these patients were followed conservatively for significant periods. Less than half of the follow-up years were considered satisfactory.

The results were approximately the same as those seen in other patients with universal colonic involvement without ileal involvement.

Four other patients with partial colonic involvement not involving the ascending colon, as demonstrated by radiologic examination, were found to have evidence of ileal disease as well.

3. *Partial Colonic Involvement.*—Ninety-six patients in this series of 148 patients had radiologic evidence of ulcerative colitis which had not involved the entire large bowel when they first came under observation. This number includes ten individuals who had partial involvement at first, which later extended to involve the entire organ, as seen on repeated radiologic examination. The results in the individuals with partial colonic involvement were fairly good. During the first three months approximately three-fourths of the patients experienced symptomatic improvement and either became asymptomatic or were

having only mild symptoms. Two patients died during the first three months while six patients were subjected to operation.

Long-term observations were carried out on 79 patients with partial involvement of the colon, who were followed for an average period of 3.1 years. The results appear in Table VIII.

TABLE VIII. LONG-TERM RESULTS IN PARTIAL COLONIC INVOLVEMENT

NUMBER OF PATIENTS	SATISFACTORY YEARS		UNSATISFACTORY YEARS	
	4	3	2	1
79	66 (27%)	61 (25%)	74 (30%)	46 (18%)

Average follow-up—3.1 years.

VOL.
4
1956

The patients were either symptom-free or experienced no more than mild symptoms during slightly more than half (52 per cent) of the follow-up years. Nine patients came to operation. There were no deaths under conservative management in this group, with the exception of one individual who died of an unrelated disease.

4. *Pseudopolyposis.*—Thirty-seven patients were observed to have evidence of pseudopolypi in the large bowel at some time during the period of study. Pseudopolyposis was more commonly observed in individuals with inflammatory involvement of the entire colon than in patients with partial colonic involvement, two-thirds of the cases with polyposis falling into the former category.

The polypoid changes were evident both on radiologic study and at proctoscopy in approximately half of these patients. It is of interest that four patients with pseudopolyposis were also discovered to have adenomatous polypi which were removed from the rectum. No instance of cancer was observed among the patients with pseudopolyposis.

TABLE IX. LONG-TERM RESULTS IN PSEUDOPOLYPOSIS

NUMBER OF PATIENTS	SATISFACTORY YEARS		UNSATISFACTORY YEARS	
	4	3	2	1
29	18 (20%)	19 (22%)	31 (35%)	20 (23%)

Average follow-up—3.0 years.

Four patients who presented pseudopolyposis were subjected to surgery without prolonged medical therapy. Three other patients came to operation later.

It was possible to appraise the clinical course under conservative management of 29 patients with pseudopolyposis who were followed for an average of 3.0 years (Table IX).

The long-term results in patients with pseudopolyposis appear to have been a little worse than those in the group as a whole. This is in conformity with the experience of others although it should perhaps be emphasized that pseudopolyposis is often compatible with a relatively satisfactory subsequent clinical course.

5. *Stricture of Bowel.*—Sixteen patients presented evidence of stricture. The constriction was located in the rectum or rectosigmoid in 13 instances. One rectal lesion proved to be malignant. Eight patients with stricture of the rectum or rectosigmoid were followed conservatively for periods exceeding a year. Three of these patients had a relatively satisfactory clinical course, while the others tended to suffer from continuous diarrhea.

Three individuals with a stricture located higher up in the large bowel were all subjected to surgery because of the uncertainty of the nature of the lesion. Two had strictures in the region of the hepatic flexure and one had a stricture of the descending colon.

6. *Anatomic Evidence of Progression of the Disease.*—Only five patients in the entire series had radiologic studies of the colon which were initially considered to be within normal limits and developed positive x-ray findings at a later date. All of the other patients presented radiologic findings initially which were interpreted as being at least suggestive of inflammatory involvement, involving the large bowel proximal to the region of the rectosigmoid.

It is of interest that during the period covered by the present study, twenty-seven additional patients with nonspecific proctitis were also observed over extended periods in the clinic. All had radiologic studies which failed to reveal evidence of the disease above the level of the rectosigmoid.

As noted earlier, ten of the patients who initially presented evidence of partial involvement of the colon were discovered on subsequent examination to have disease of the entire large bowel.

It appears that convincing evidence of a spread in the anatomic extent of the disease is encountered in a minority of patients with ulcerative colitis whose clinical course is observed over a period of years.

DURATION OF DISEASE

Various observers have reported that ulcerative colitis tends to be more severe during the first year or two of the disease[8] and that the mortality rate is higher during this period. The significance of a short preceding history was evaluated in this series in 38 patients in whom the duration of bowel symptoms was less than one year. Twenty-four of these patients had partial involvement of the large bowel, while in fourteen total involvement was demonstrated by radiologic examination.

The immediate results were not significantly different from those observed in the entire series. Approximately two-thirds of the patients became either symptom-free or the symptoms became mild in severity during the first three months of observation. Two patients died while three other patients underwent early surgery.

Twenty-eight of the patients with a short preceding history were followed medically for over six months, the average total follow-up being 3.2 years. As

seen in Table X, the long-term results are similar to those observed for the whole series. Approximately half of the follow-up years were relatively satisfactory.

Two patients came to operation subsequently. Eight patients had x-ray evidence of progression from partial to complete colonic involvement during the follow-up period.

TABLE X. LONG-TERM RESULTS IN EARLY CASES

NUMBER OF PATIENTS	SATISFACTORY YEARS		UNSATISFACTORY YEARS	
	4	3	2	1
28	23 (26%)	21 (24%)	18 (20%)	27 (30%)

Average follow-up—3.2 years.

OL.
4
956

These data do not suggest that the duration of the disease prior to observation bears a striking relation to the immediate or long-term outlook. The severity of subsequent symptoms in this group of individuals, who had had the disease for less than one year, did not differ significantly from the group as a whole. However, the incidence of incapacitating symptoms was somewhat greater in this small group than in the general average of patients with ulcerative colitis.

OTHER COMPLICATIONS

With the exception of fistula in ano, most of the complications of chronic ulcerative colitis were not common in this series of cases. Only one patient developed a perforation of the colon and one had massive bleeding. Extensive necrotic skin lesions were encountered in three individuals. There was one case of optic neuritis and one of choroiditis. The coexistence of cirrhosis of the liver was established in three patients.

Seventeen patients suffered from chronic arthritic symptoms. The pattern varied from migratory joint pains to a frank clinical picture of rheumatoid arthritis. Two patients had Marie Strümpell arthritis.

This study casts no light on the risk of development of cancer of the large bowel in patients with ulcerative colitis. Only one individual in the series developed carcinoma. The increased risk of this complication of ulcerative colitis has been demonstrated by others in larger series of cases.[9]

USE OF CORTISONE OR CORTICOTROPIN

Cortisone or corticotropin were, with a few exceptions, employed only when patients were seriously ill and had failed to respond to other measures. These preparations were used in the treatment of 21 patients in this series for periods varying from one week to several months. The dosage of cortisone was usually begun at 300 mg. and reduced rapidly to 100 mg. or less per day. Corticotropin was used in comparable dosage in a few patients.

The apparent effect of hormone therapy was varied. Nine patients went into early complete—or nearly complete—symptomatic remission. The effect of treatment was less striking or disappointing in the remaining patients. One

patient sustained a perforation of the large bowel, one died of a pulmonary embolus, and another individual had a transitory psychosis. Six of the 21 patients who received cortisone or corticotropin were ultimately subjected to surgical intervention.

Eleven patients were followed conservatively after the conclusion of hormone therapy for periods which varied from six months to two years. The result was approximately that which would be anticipated in the more severe forms of ulcerative colitis. The patients experienced moderate to severe exacerbations during approximately two-thirds of the follow-up years.

DISCUSSION

The immediate results of medical management of ulcerative colitis were fairly satisfactory in two-thirds of the patients in the present series, the symptoms becoming mild or disappearing. This is in general accord with the results reported in the past fifteen years by various investigators, using supportive therapy and various sulfa drugs or antibiotics. Block and Pollard[10] summarized the reported experiences of many authors with sulfonamides and stated that benefit was noted in approximately 70 per cent of patients after relatively short periods of observation. Similar experiences have been recorded with the use of Aureomycin, Chloromycetin, and other antibiotics.[11,12]

The long-term outlook in ulcerative colitis in the present series is also similar to that reported by others in both the earlier and the more recent literature.[1-6,13,14] Patients who experience complete and sustained remission were exceptional. In this respect there has been no improvement in the results of medical treatment of ulcerative colitis in the past twenty years. However, the mortality rate from ulcerative colitis has probably decreased appreciably in the past twenty years. The mortality among the medically treated patients in this series was 6.1 per cent. The early experiences reported by Hardy and Bulmer[4] and Hurst[5] recorded mortality rates for all patients of 33 to 50 per cent. However, these data were based upon surveys which extended over a greater period of years. The difference is probably attributable, in part, to improved methods of medical treatment, even though the methods be nonspecific.

A clinical picture either of severe constitutional symptoms over a period of weeks under medical management or of unremitting, although less severe, colonic symptoms over a period of months, presents grave problems. It has long been noted that patients with a so-called toxic clinical picture have a grave outlook. This was found to be particularly true in the present series in those patients who exhibited a fever exceeding 103° F. when the fever lasted two weeks or more. Such patients often were considered to warrant early operation, or eventually required it later.

The patient who does not have associated severe constitutional symptoms but who, nevertheless, continues to suffer from moderate to severe diarrhea for many months, likewise had a relatively poor outlook in the present series. This again is in accord with the experiences of others. Hern[3] in 1931, for example, emphasized the relative gravity of the chronic type of ulcerative colitis devoid of periods of remission in contrast to the so-called intermittent type. The reason

why failure to remit under treatment is a poor prognostic sign is not clear in most instances but presumably reflects continuing activity of the causative factors in the disease.

In the present study the most favorable long-term results were observed in those patients who demonstrated what appeared grossly to be a complete remission in the inflammatory process as viewed through the sigmoidoscope. This may take place many months or even years after symptoms have entirely disappeared. Complete proctoscopic remission was noted in only a minority of the patients in the present series. A higher proportion might have been recorded if this examination had been performed with greater frequency after the symptoms had subsided. Further data concerning the rate of proctoscopic remission in this disease would be of value and should be useful for the evaluation of newly proposed therapeutic measures.

OL.
4
956

The severity and extent of anatomic involvement in ulcerative colitis have long been regarded as important in prognosis, although opinion has varied somewhat as to how much significance to attach to these factors.[6,7] In this series, total involvement of the large bowel regardless of severity was characterized by a long-term outlook which was distinctly poorer than the average, two-thirds of the follow-up years being unsatisfactory in such patients.

Ileal involvement, shown on radiologic examination, in addition to universal colonic disease, appeared to have little or no additional effect on prognosis. The extent of ileal involvement as seen by x-ray was usually limited to a small segment of the ileum proximal to the ileocecal valve, and extensive anatomic disease of the small bowel was unusual in this series. Current studies of associated functional disorders of the small intestine may throw further light on this phase of the problem.

Other anatomic factors may be important in prognosis. Stricture of the rectum in this series was usually associated with persistent intractable diarrhea, although exceptions were encountered. On the other hand, the occurrence of pseudopolyposis did not correlate with a very poor long-term prognosis in the present series. It occurred more often in the patients with total involvement of the colon than in patients with partial involvement. This manifestation of the repair process may be regarded simply as a sequel of extensive tissue injury.

Our experience with cortisone and corticotropin in the treatment of the seriously ill patient has been somewhat disappointing. The administration of these substances was followed in some patients by a prompt improvement in constitutional symptoms, including a fall in temperature, improvement in appetite and sense of well-being, as well as a decrease in diarrhea. In other patients the use of these hormones appeared to have no beneficial effect. Our limited use of hormone therapy in selected severely ill patients is not suitable for comparison with the recent studies of English investigators,[15,16] who found that cortisone was of definite benefit both symptomatically and objectively in an essentially unselected series of patients with this disease.

CONCLUSIONS

1. The early and late results of medical management of a series of patients with ulcerative colitis have been summarized. The following generalizations seem warranted: (a) Supportive therapy yields relatively satisfactory early

results in two-thirds of patients. (b) Over long periods, approximately one out of two follow-up years may be regarded as relatively satisfactory in this disease.

2. Certain clinical features of ulcerative colitis aid in prognosis, including the following: (a) A lack of symptomatic response within the first three months of treatment was correlated with generally unsatisfactory long-term results. (b) Complete proctoscopic remission, within the limits defined, usually is associated with an excellent subsequent prognosis. (c) Extensive colonic involvement is usually, but not invariably, associated with a poorer prognosis.

3. Individuals with severe constitutional symptoms, including a high temperature persisting over two weeks under conventional medical treatment, have usually required surgical intervention.

REFERENCES

1. Bargen, J. A.: Chronic Ulcerative Colitis, Springfield, Ill., 1951, Charles C Thomas.
2. Crohn, B. B., and Rosenberg, H.: The Sigmoidoscopic Picture of Chronic Ulcerative Colitis (Non-specific), Am. J. M. Sc. **170**:220, 1925.
3. Hern, J. R. B.: Ulcerative Colitis, Guy's Hosp. Rep. **11**:322, 1931.
4. Hardy, T. L., and Bulmer, E.: Ulcerative Colitis, A Survey of Ninety-Five Cases, Brit. M. J. **2**:812, 1933.
5. Hurst, A. F.: Prognosis of Ulcerative Colitis, Lancet **2**:1194, 1935.
6. Kiefer, E. D.: Clinical Results in the Medical Treatment of Chronic Ulcerative Colitis, Am. J. Digest. Dis. & Nutrition **3**:56, 1936.
7. Ricketts, W. E., Kirsner, J. B., and Palmer, W. L.: Chronic Non-specific Ulcerative Colitis —A Roentgenologic Study of Its Course, Gastroenterology **10**:1, 1948.
8. Rice-Oxley, J. M., and Truelove, S.: Ulcerative Colitis—Course and Prognosis, Lancet **1**:663, 1950.
9. Sloan, W. P., Bargen, J. A., and Baggenstoss, A. H.: Local Complications of Chronic Ulcerative Colitis Based on the Study of 2000 Cases, Proc. Staff Meet., Mayo Clin. **25**:240, 1950.
10. Block, M., and Pollard, H. M.: Chemotherapy and Antibiotics in Chronic Ulcerative Colitis, Gastroenterology **10**:46, 1948.
11. Berkowitz, Z. T.: Chloromycetin Therapy of Chronic Ulcerative Colitis—a One Year Study, Gastroenterology **16**:19, 1950.
12. Wright, L. T., Strax, S., Marks, J. A.: Treatment of Non-specific Ulcerative Colitis With Aureomycin, Ann. West. Med. & Surg. **4**:717, 1950.
13. Kapel, O.: Medical and Modern Surgical Treatment of Chronic Ulcerative Colitis, Acta med. scandinav. **138**:328, 1950.
14. Svartz, N.: The Pathogenesis and Treatment of Ulcerative Colitis, Acta med. scandinav. **141**:172, 1951.
15. Truelove, S. C., Witts, L. J., Tunbridge, R. E., Watkinson, G., Jones, F. A., Doll, R., Hardy, T. L., St. Johnston, C. R., Card, W. I., Wilson, M., and Taylor, J.: Cortisone in Ulcerative Colitis. Preliminary Report on a Therapeutic Trial, Brit. M. J. **11**:375, 1954.
16. Jones, F. A.: Ulcerative Colitis, California Med. **82**:430, 1955.

VOL 4 195

CARBOHYDRATE METABOLISM AND NUTRITIONAL STATE IN MULTIPLE SCLEROSIS

H. DROLLER, M.D., M.R.C.P., AND I. J. N. POWELL, B.SC.

LEEDS, ENGLAND

From the Geriatric Unit, St. James's Hospital

(Received for publication Dec. 20, 1955.)

RECENT reports have suggested a disturbance of carbohydrate metabolism in multiple sclerosis.[1-3] These reports are based on the observation of glucose tolerance tests and pyruvic acid levels. The latter were considered particularly important since they implied a thiamine deficiency. The diphosphate of thiamine (cocarboxylase) is required for the decarboxylation of pyruvate, and in its absence, glucose oxidation stops at pyruvic acid level. Pyruvate then accumulates in tissues and blood.[4]

The present study has been designed to investigate the significance of disturbed carbohydrate metabolism and thiamine deficiency in multiple sclerosis by biochemical investigation and clinical observation.

METHODS

Blood sugar estimations were carried out by the Hagedorn and Jensen[5] method, using the protein precipitation of Somogyi.[6]

Blood for pyruvic acid estimation was collected in Long's[7] mixture and refrigerated until estimated according to Friedemann and Haugen.[8]

The patients had been fasting for at least twelve hours. Fifty grams of glucose dissolved in 400 ml. of water was given orally after fasting blood samples had been withdrawn, and blood sugar and pyruvic acid were estimated at half-hour intervals for two hours.[9]

All patients subjected to the tests had clinically fully established multiple sclerosis. The details appear in Table I.

Patients are referred to throughout the various tests by the same number. They were resident in the long-stay annex of the Geriatric Unit. The nature of the investigation was explained in simple terms and cooperation was gladly given by ten patients.

We have incomplete data on several other patients who refused to cooperate for various reasons. We shall refer to them only in regard to the mean fasting pyruvic acid level of blood.

THE FIRST GLUCOSE TOLERANCE TESTS

Three types of glucose tolerance curves were observed (Table II): a flat curve, a curve with steep rise and return to the initial level within the second hour (lag curve), and a diabetic type of curve (Table II). The fasting level was abnormally high in 2 and abnormally low in 2 patients. The four flat curves (Nos. 1, 2, 4, and 8) suggest an excessive intake of carbohydrate or some mal-

TABLE I. CLINICAL DETAILS OF PATIENTS

PAT. NO.	(YR.) AGE	SEX	DURATION (YR.)	PARALYSIS OF HANDS	ATAXIA OF HANDS	BEDFAST	CHAIR-FAST	AMBU-LANT
1	53	F	12	—	marked	−	+	−
2	59	M	17	—	marked	−	−	+
3	43	M	11	—	Slight	−	+	−
4	35	F	13	Weakness	Marked	+	−	−
5	39	F	16	—	Moderate	−	+	−
6	50	M	6	—	Moderate	+	−	−
7	52	M	7	Weakness	Slight	+	−	−
8	49	M	7	—	Marked	−	−	+
9	61	F	18	—	Moderate	—	+	−
10	55	F	21	Paralyzed	Slight	+	−	−

absorption. As normal, we considered a blood sugar curve which rose after the ingestion of 50 Gm. of glucose in 300 ml. of water from not more than 110 mg. per cent to a maximum of 200 mg. per cent and fell within two hours to 110 mg. per cent or less.[9,10]

TABLE II. GLUCOSE TOLERANCE CURVES AT THE START OF THE EXPERIMENT

	FAST	½ HOUR	1 HOUR	1½ HOURS	2 HOURS	COMMENT
1	103	104	111	114	88	Flat
2	86	93	88	67	65	Flat
3	100	152	158	176	182	Diabetic
4	60	83	106	106	110	Flat
5	124	202	124	110	110	Lag
6	121	149	152	155	160	Diabetic
7	89	120	133	164	138	Diabetic
8	64	122	115	89	77	Flat
9	138	251	213	195	155	Diabetic
10	61	138	153	182	217	Diabetic

DIETARY INVESTIGATION

None of these patients had been recognized clinically as diabetic and the urine of only one patient was known to contain sugar (No. 10). It was decided to submit the patients to a dietary investigation in order to assess the influence of the diet on the blood sugar curve. The patients had free choice of foods, but both offered and rejected food was weighed. Daily food intake charts were kept. For estimation of caloric values and vitamins we used the tables of Jolliffe.[11] The dietary intake was estimated for seven days in each case and the calories consumed per patient per day shown in Table III are the calories consumed during this time, divided by seven.

These diets appear to be just compatible with a life which requires almost no muscular exertion. It will be noted that some of the diets were high in carbohydrates and all had a sufficiency of protein and fat.

Thiamine and niacin requirements were covered, though on the low side.[12-14]

The caloric requirements of chronic invalids are not at all well known. Jolliffe[11] quotes the following for sedentary occupations: man 2,400 calories, woman 2,000 calories.

Both male and female patients ate less than Jolliffe's estimates. In parallel dietary investigations on Parkinsonians, we found a mean caloric intake in three men of 2,300 and in three women of 2,200 calories when spoon-fed. Self-feeding Parkinsonians ate more than the average—2,500 to 3,000 calories (taking spilled food into account).

TABLE III. CALORIC INTAKE OF MULTIPLE SCLEROSIS PATIENTS PER DAY

		CAL.	PROTEIN (GM.)	FAT (GM.)	C. H. (GM.)	THIAMINE (MCG)	NIACIN (MG)
1	c	2,135	64	94	263	710	9.5
2	a	3,137	101	119	426	1,077	13.7
3	b	2,093	61	103	240	639	9.2
4	b			deterioration			
5	c	1,261	39	72	115	514	4.8
6	b	1,984	55	96	217	568	11.6
7	b	1,398	44	60	174	502	4.5
8	a	2,263	73	98	301	785	10.3
9	c	1,405	35	62	180	462	3.5
10	b	1,702	50	73	220	612	6.3

a = ambulant; b = bedfast; C = chair bound.

Patient 4 lived almost exclusively on 5 per cent glucose drinks which she sucked up through a rubber tube. She was too ataxic to feed herself and later on deteriorated to such an extent that the dietary investigation could not be carried out.

The mean for the two sexes separately is as follows:

		Male	*Female*
Calories		2,170	1,627
Protein	Gm.	66	47
Fat	Gm.	95	75
Carbohydrate	Gm.	271	194
Thiamine	Mg.	714	574
Niacin	Mg.	9.8	6.2

DIET AND GLUCOSE TOLERANCE TESTS

The effect of diet on glucose tolerance tests is well known.[15-17] The influence of starvation on glucose tolerance curves has been investigated by many authors[18-21] and the curves which we obtained in our multiple sclerosis patients were of a type compatible with undernutrition. Some authors[18,20] saw more commonly diabetic curves, whereas Bose[19] found flat curves after oral ingestion of glucose and diabetic curves after intravenous glucose injection. Sherlock and Walshe[22] found normal glucose tolerance curves in the undernourished people of the Rhineland.

Additional glucose feeds: After the preliminary glucose tolerance tests and the dietary investigation had revealed a caloric deficiency, it was decided to submit the patients to feeding with extra glucose.

The main purpose was to discover whether extra loading with glucose, given for a prolonged period, could be adequately dealt with by these patients.

Glucose was given, not as a food but as a medicine, to ensure that it was readily taken by the patients in daily doses of 100 Gm. suitably flavored and

diluted with water. Glucose feeding continued for three weeks after which the glucose tolerance tests were repeated (Table IV).

The results show that the tolerance test curves did return to a more normal shape in all except one who showed the curve of frank diabetes. One man (No. 3) and three women (Nos. 4, 5, and 9) showed, after glucose feeding, some delay of return (lag curves). The diabetic patient (No. 10) was excluded from further participation in the experiment.

TABLE IV. GLUCOSE TOLERANCE TESTS AFTER GLUCOSE FEEDING

	FAST	½ HOUR	1 HOUR	1½ HOURS	2 HOURS	COMMENT ON FIRST G.T.T.	COMMENT ON SECOND G.T.T.
1	95	170	152	114	80	Flat	Normal
2	78	138	147	67	80	Flat	Normal
3	92	141	177	177	141	Diabetic type	Lag curve
4	108	146	172	147	124	Flat	Lag curve
5	96	160	187	199	160	Lag	Diabetic
6	100	161	130	100	97	Diabetic	Normal
7	112	116	148	124	107	Diabetic	Normal
8	101	122	147	112	101	Flat	Normal
9	106	191	226	157	127	Diabetic	Lag curve
10	200	255	260	284	290	Diabetic	Diabetic

THIAMINE AND COCARBOXYLASE AND THEIR INFLUENCE ON THE GLUCOSE TOLERANCE TEST

Although Lawrence and Oakley[23] did not observe any effect of thiamine on the insulin requirements of diabetics, there is some evidence that thiamine may influence blood sugar level and glycogen storage.[24-27] It has been asserted that thiamine diphosphate or cocarboxylase was more insulin-saving in diabetic coma than thiamine.[25-27] Since thiamine becomes immediately phosphorylated when added to many animal tissues,[28] it appears doubtful whether phosphorylated vitamin B could differ in its effect on carbohydrate metabolism from ordinary thiamine.

A supply of cocarboxylase was available to us and it was decided to submit patients after the period of glucose feeding to daily intramuscular injections of 50 mg. of cocarboxylase for periods of 14 days, after which the glucose tolerance tests were repeated. Following this, 50 mg. of thiamine was injected daily for 14 days by the same route and glucose tolerance tests repeated once again (Tables V and VI).

TABLE V. GLUCOSE TOLERANCE TESTS AFTER THIAMINE TREATMENT

	FAST	½ HOUR	1 HOUR	1½ HOURS	2 HOURS	
1	92	190	184	141	70	Normal
2	78	138	150	122	76	Normal
3	70	132	132	121	79	Normal
4	90	140	122	115	95	Normal
5	74	150	152	122	106	Lag curve
6	110	132	148	115	100	Normal
7	116	141	148	161	111	Normal
8	83	133	151	101	81	Normal
9	113	131	173	141	131	Lag curve

Both thiamine and cocarboxylase tended to normalize the glucose tolerance curves of the patients with high blood sugar curves. The results do not suggest that the cocarboxylase is superior in its action on the glucose tolerance curve compared with thiamine. Indeed thiamine and cocarboxylase injections may have acted in this favorable way, simply by stimulating the appetite, thereby causing a more adequate carbohydrate intake.

TABLE VI. GLUCOSE TOLERANCE TESTS AFTER COCARBOXYLASE TREATMENT

	FAST	½ HOUR	1 HOUR	1½ HOURS	2 HOURS	
1	92	180	165	120	90	Normal
2	80	145	130	115	80	Normal
3	84	110	127	101	70	Normal
4	100	148	150	110	97	Normal
5	103	120	140	128	122	Mild diabetic
6	86	150	104	82	72	Normal
7	110	169	140	139	100	Normal
8	82	135	150	109	77	Normal
9	100	152	175	132	130	Lag curve

OL.
4
956

THE BEHAVIOR OF PYRUVIC ACID IN MULTIPLE SCLEROSIS FOLLOWING THE INGESTION OF GLUCOSE

It has been stated that multiple sclerosis patients show a high fasting level of blood pyruvic acid.[3] The normal fasting pyruvic acid level found at our laboratory was 0.6 ± 0.4 mg. per cent. This conforms to acknowledged standards.[29] According to Klein,[29] the pyruvic acid level does not change with age or sex. In a series of 15 undernourished persons not suffering from heart disease or diabetes, we found a fasting pyruvic acid level of 1.13 ± 0.5 mg. per cent. In the multiple sclerosis patients the pyruvic acid level in the fasting state (prior to feeding periods) was 0.83 ± 0.4 mg. per cent.

Although the fasting level in patients with multiple sclerosis is lower than that of undernourished persons, the smallness of the sample does not permit us to state whether this is statistically significant.

Although there was no obvious clinical or biochemical evidence of thiamine deficiency in our patients, it was decided to follow the pyruvic acid simultaneously with the glucose tolerance test.

Turnock and Welbourne[30] pointed out that thiamine deficiency can be presumed if the pyruvic acid level fails to return to the basal level three hours after the ingestion of 100 Gm. of glucose, even if peak and fasting levels are normal. They also stated that, after treatment with thiamine, previously abnormal curves returned to normal.

Pyruvic acid estimations carried out simultaneously with oral glucose tolerance tests (Table VII) showed no evidence of thiamine deficiency. Although we used a two-hour test and only 50 Gm. of glucose, blood pyruvic acid had returned to the fasting level within that period in six patients in the first (pre-glucose feeding) test.

After glucose feeding all except Patient 2 showed pyruvic acid levels at two hours of the same order as the fasting level. This man ingested the largest amount of carbohydrates of all the patients, and he might have become temporarily thiamine deficient though he had 1,077 micrograms of thiamine in his

diet. If thiamine depletion had been present in the others, the additional glucose feeding should have accentuated it.[31]

After 2 weeks of injections of cocarboxylase followed by 2 weeks of thiamine injections, no marked change was observed in the pyruvic acid level. One may, therefore, conclude that none of our multiple sclerosis patients was thiamine-depleted.

TABLE VII. PYRUVIC ACID (MILLIGRAMS PER CENT)

		FAST	½ HOUR	1 HOUR	1½ HOURS	2 HOURS
1	Before	0.72	—	—	—	—
	Glucose	0.60	0.83	0.37	insuff.	insuff.
	Cocarboxylase	0.54	0.65	0.87	0.63	0.65
	Thiamine	0.98	1.65	2.20	1.40	1.33
2	Before	0.82	0.98	1.10	0.88	0.42
	Glucose	0.43	0.39	0.56	0.56	0.72
	Cocarboxylase			not performed		
	Thiamine	0.68	0.77	1.49	0.66	0.66
3	Before	0.40	0.70	0.53	0.40	0.38
	Glucose	0.55	0.71	0.62	0.84	0.42
	Cocarboxylase	0.32	0.42	0.33	0.42	0.45
	Thiamine	0.33	0.85	0.47	0.45	0.23
4	Before	1.05	1.45	0.97	0.85	0.95
	Glucose	1.19	1.66	1.86	1.21	0.96
	Cocarboxylase			not performed		
	Thiamine	0.9	1.1	0.8	0.8	0.76
5	Before	0.75		not performed		
	Glucose	0.54	0.48	0.58	0.45	0.45
	Cocarboxylase	0.59	1.05	0.67	0.48	0.13
	Thiamine	0.42	0.62	0.69	0.57	0.40
6	Before	0.80	1.20	1.15	1.00	0.72
	Glucose	0.95	1.30	1.00	0.88	0.90
	Cocarboxylase	0.80	1.00	0.58	0.37	0.45
	Thiamine	0.85	1.10	0.73	0.42	0.88
7	Before	0.22	0.27	0.22	0.16	0.16
	Glucose	0.80	1.08	1.19	0.63	0.67
	Cocarboxylase			not performed		
	Thiamine	0.27	0.50	0.36	0.27	0.27
8	Before	0.74		not performed		
	Glucose	0.60	2.10	0.53	0.44	0.44
	Cocarboxylase	0.39	1.06	0.44	0.39	0.19
	Thiamine	1.08	1.51	1.28	0.60	1.06
9	Before	0.59	0.97	0.75	0.71	0.46
	Glucose	0.87	1.26	2.32	1.61	0.42
	Cocarboxylase	0.88	1.10	2.35	1.19	0.69
	Thiamine	0.47	0.57	0.48	0.33	0.20

CLINICAL OBSERVATIONS

All patients remained resident in the Geriatric Unit throughout the various procedures. Weekly neurologic examinations were carried out without consideration of the biochemical program. The clinical interview included enquiry

into general well-being, general mobility, ability to turn over in bed, and feeding habits. Neurologic examination included tests for ataxia, intention tremor, state of tendon reflexes, and increase or diminution of spasticity.

The results were disappointing. No major remission was observed throughout the whole time the patients were under observation, a minimum of eight weeks. Since the conclusion of the program one patient has died, the others have been kept under supervision for a further twelve months. No major remissions were noted in any of them.

All patients stated at the conclusion of the tests that they felt better. Two (Nos. 4 and 5) had relapses during cocarboxylase therapy with increase in spasticity and weakness. One (No. 6) developed urinary retention during thiamine therapy. Two patients had minor improvements during thiamine therapy with less ataxia and diminished spasticity (Nos. 5 and 7). At the conclusion of the investigation two were definitely worse, the others unchanged.

The picture is that classically described in multiple sclerosis: minor improvements and deteriorations vary with the frequency of clinical appraisal. Major relapses and remissions are comparatively few and far between.

DISCUSSION

There is at the moment no rational therapy for multiple sclerosis. Markees[25,26,27,32,33] investigated pyruvic acid metabolism and postulated a cocarboxylase deficiency as the basic disturbance in the cell metabolism of multiple sclerosis. Following this Ervenich[2,3] investigated sixty-seven cases of this disease and found high pyruvic acid levels and abnormal glucose tolerance curves. He treated his patients with cocarboxylase injections and claimed improvement in 70 per cent of his cases. As soon as cocarboxylase was withdrawn the patients relapsed.

Such therapeutic success could not be reproduced in the present series.

We could rectify the alleged inherent abnormality of the carbohydrate metabolism by the oral administration of glucose. The concomitant improvement of glucose tolerance tests and fasting pyruvic acid levels signified that our patients' nutritional state had been suboptimal.

This raises the question of the optimal nutrition of chronically ill people and in particular of multiple sclerosis patients. We are not aware of any work on the diets of patients suffering from multiple sclerosis.

Our patients were offered a diet containing about 3,000 calories daily, but in actual fact they ate less. We do not know the optimum caloric value of a multiple sclerosis patient's diet, nor do we know whether minor dietary deficiencies influence the clinical picture.

Clearly our patients had starved themselves from fear of spilling food. Later in the disease, weakness and ataxia made adequate self-feeding impossible. The euphoric and polite sufferer from multiple sclerosis in institutional care seems incapable of making his nutritional needs known, or, to avoid conflict, goes without food.

The small caloric intake is compatible with an existence requiring little muscular effort. A group who subsisted on a similar low food intake was de-

scribed by Bransby and Osborne.[34] They observed a caloric intake of 2,170 in elderly men and 1,700 calories in women of the same age. It is obvious that the offering of good, well-balanced diets to chronic invalids is not enough unless one makes sure they eat them.

Two cases with abnormal glucose tolerance curves (Nos. 5 and 9) cannot be considered true diabetics because the fasting levels of blood sugar were normal and the peak levels were below 180 mg. per cent. The delay in return to the original level is probably due to the nutritional state of the patients.

One female patient was found to have diabetes mellitus (No. 10). Here the diabetes was masked by the low nutritional state but it became quickly apparent after glucose feeding. Kendall[35] states that diabetes is rare in female patients suffering from multiple sclerosis. It may, however, signify that many multiple sclerosis patients live on the borderline of starvation, and the low caloric intake masks the diabetes. Diabetes mellitus is rapidly increasing[36] and it seems probable that the two conditions coexist. There is no evidence that carbohydrate metabolism is at fault in multiple sclerosis.

It is of special importance to insist on the establishment of satisfactory basic nutritional conditions before embarking on glucose tolerance tests in chronic invalids. A single glucose tolerance test may not be sufficient and a further test should be done after a period of loading with carbohydrate or glucose. We emphasize this point because of our experience with the first series of tests which seemed to provide impressive evidence in favor of a disturbance of carbohydrate metabolism. After a satisfactory food intake had been established, the glucose tolerance curves became normal.

From the dietary investigation and the pyruvic acid curves, no evidence of a thiamine deficiency could be deduced. Preparations of thiamine and phosphorylated cocarboxylase showed no effect on the clinical state of multiple sclerosis patients. None of the claims made by Ervenich[2,3] could be substantiated.

SUMMARY

An investigation was undertaken into the nutrition and carbohydrate metabolism of ten patients suffering from multiple sclerosis who subsisted on diets of their own choice. The general level of nutrition was poor, intake of fat and protein was low in both sexes, and the carbohydrate intake was low in women. Thiamine and niacin of the diet appeared adequate.

The observed patterns of abnormal glucose tolerance tests corresponded to those seen in starvation and could be rectified by glucose feeding.

One case of diabetes mellitus became apparent after glucose feeding.

There was no biochemical evidence of thiamine deficiency in these patients subsisting on poor diets. There was evidence of temporary thiamine deficiency in one patient after extra loading with glucose.

Glucose tolerance curves, following the administration of thiamine and cocarboxylase, respectively, were of a similar pattern.

Cocarboxylase and thiamine were found to be of no therapeutic value.

We wish to thank Professor R. E. Tunbridge for his advice, and Miss M. V. Grogan, Dietitian to the Hospital, for her help with the investigation of the diets.

We gratefully acknowledge a supply of cocarboxylase by Roche Products, Ltd.

REFERENCES

1. Jones, H. H., Jones, H. H., Jr., and Bunch, L. D.: Biochemical Studies in Multiple Sclerosis, Ann. Int. Med. **38**:831, 1950.
2. Ervenich, P.: Neue Gesichtspunkte zur Behändlung der Multiplen Sklerose, Äerztliche Praxis **4**:2, 1952.
3. Ervenich, P.: Neue Gesichtspunkte zur Behändlung und Aetiologie der Multiplen Sklerose, Ärztl. Forsch. **7**:55. 1953.
4. Platt, B. S., and Lu, G. D.: Studies of Metabolism of Pyruvic Acid in Normal and Vitamin B Deficient States, Biochem. J. **33**:1525, 1939.
5. Hagedorn, H. L., and Jensen, B. N.: Blutzuckerbestimmung, Biochem. Ztschr. **135**:4, 1923
6. Somogyi, M.: Determination of Blood Sugar, J. Biol. Chem. **160**:69, 1945.
7. Long, C.: Stabilisation and Estimation of Pyruvic Acid in Blood Samples, Biochem. J. **38**:447, 1944.
8. Friedemann, T. E., and Haugen, G. E.: Determination of Ketoacids in Blood and Urine, J. Biol. Chem. **147**:415, 1943.
9. Evensen, O. K.: Alimentary Hypoglycaemia, Acta. med. scandinav. Suppl. **126**: 1942.
10. Siosalo, P.: On the Blood Sugar Curve of Healthy Persons, Acta. med. scandinav. **34**:184, 1930.
11. Jolliffe, N. (Ed): Clinical Nutrition, New York, 1950, Paul B. Hoeber, Inc.
12. Cowgill, G. R.: Vitamin B Requirements in Man, New York, 1943.
13. Kodicek, E.: Minimal Requirements of Nicotinic Acid, Lancet **1**:380, 1942.
14. Harris, L. J., and Long, P. C.: Vitamins in Human Nutrition, Lancet 886, 1936.
15. Sweeney, J. S. C.: Dietary Factors That Influence the Dextrose Tolerance Test, Arch. Int. Med. **40**:818, 1927.
16. Sweeney, J. S. C.: A Comparison of the Effect of General Diets and of Standard Diets on Tolerance for Dextrose, Arch. Int. Med. **42**:872, 1928.
17. Himsworth, H.: Mechanism of Diabetes Mellitus, Lancet **2**:1, 1939.
18. Robinson, G. W., Jr., Shelton, P., and Smith, F. V., Jr.: Importance of Dextrose Tolerance Test in Marginal Malnutrition, Arch. Int. Med. **68**:945, 1941.
19. Bose, J. P., De, U. N., and Mukerjee, P.: Preliminary Study of Biochemical Changes in Starvation Cases, Ind. J. Med. Res. **34**:143, 1946.
20. Klotzbuecher, E.: Klinische Beobachtungen bei der Oedemkrankheit, Klin. Wchnschr. **26**:289, 1948.
21. Helweg, Larsen, Hoffmeyer, P. H., Kieler, J., Thaysen, E. Hess, Thaysen, J. Hess, Thygessen, P., and Hartel, W. M.: Famine Disease in German Concentration Camps, Complications and Sequels, Acta. med. scandinav. Suppl. 142 1952.
22. Sherlock, S., and Walshe, V.: Disturbance of Liver Function in Studies of Under-nutrition, Wuppertal, 1946-1949, Spec. Rep. Med. Res. Coun. No. 275, London, 1951.
23. Lawrence, R. D., and Oakley, W. G.: Letter to the Editor, Vit. B_1 and Insulin, Brit. M.J. 1227, 1938.
24. Morelli, A., and d'Ambrosio, L.: Vitamina B_1 e Metabolismo degli Idrati di Carboni, Boll. Soc. Ital. Biol. Spec. **14**:401, 1939.
25. Markees, S., and Meyer, W.: Zum Stoffwechsel Problem der Brenztraubensaüre, Schweiz. med. Wchnschr. **79**:931, 1949.
26. Markees, S.: Verkommen und Bedeutung von Stoerungen im Brenztraubensaüre-Stoffwechsel, Helvet. med. acta. Series A **17**:537, 1950.
27. Markees, S.: Stoffwechsel Grundlagen und Indikationsgebiete fuer eine Therapie mit Kokarboxylase, Deutsche med. Wchnschr. **78**:971, 1953.
28. Ochoa, S.: Enzymic Synthesis of Cocarboxylase in Animal Tissues, Biochem. J. **33**:1262, 1939.
29. Klein, J. R., Perlzweig, W. R., and Handler, P.: Determination of Nicotinic Acid in Blood Cells and Plasma, J. Biol. Chem. **145**:27, 1942.
30. Turnock, D. M., and Welbourne, R. B.: Effects of Glucose on Blood Pyruvic Acid Levels in Mild Thiamine Deficiencies, J. Lab. & Clin. Med. **42**:261, 1953.
31. Arnold, A., and Elvehjem, L. A.: Influence of Composition of Diet on Thiamin Requirements of Dogs, Am. J. Physiol. **126**:289, 1939.
32. Markees, S.: Zum Problem der Brenztraubensaüre in vivo, Experientia **4**:31, 1948.
33. Markees, S.: Zum Problem der Brenztraubensaüre in vivo, ibid. **4**:195, 1948.
34. Bransby, E. R., and Osborne, B.: A Social and Food Survey of the Elderly Living Alone or as Married Couples, Brit. J. Nutrition **7**:160, 1953.
35. Kendall, V.: Disseminated Sclerosis and Diabetes Mellitus, Letter to the Editor, Lancet **2**:202, 1953.
36. Tunbridge, R. E.: Socio-medical Aspects of Diabetes Mellitus, Lancet **1**:893, 1953.

REHABILITATION OF THE ELDERLY AMPUTEE

BRUCE B. GRYNBAUM, M.D.,* NEW YORK, N. Y., AND EDWARD E. GORDON, M.D.,** CHICAGO, ILL.

(Received for publication Jan. 15, 1956.)

IN THE field of chronic illness, one of the most important problems that confronts the rehabilitation team is the elderly lower limb amputee. Vas cular impairment, which occurs with increasing frequency in the older age group, implicates the legs with relative frequency and rarely involves the arms. The loss of a leg demands replacement if it is at all feasible. With vascular changes in the legs, there may occur similar deficiencies in other organs, particularly brain, heart, and kidney, which impair physical capacity. These factors may complicate the course of management subsequent to the loss of the limb. Also, it is generally true that motivation tends to decline in the older age group because of defeatism and apathy in the face of catastrophe.

The major pathologic event leading to amputation is gangrene due to occlusion of the vessels by thrombosis on the basis of arteriosclerosis. This disease process is usually more pronounced in the presence of diabetes, which is prevalent in the elderly group. Another mechanism of vascular occlusion is embolism, frequently secondary to abnormal cardiac rhythms or a mural thrombus. Less frequent as a cause of gangrene is Buerger's disease. There are several rarer causes of amputation in the elderly such as frostbite and malignancies of bone and soft tissue. In addition, although trauma figures largely in the younger age group, the elderly are not exempt from violent injury in our technologic society. Osteomyelitis, nonunion of fractures, and syphilitic complications have largely yielded to newer medical and surgical techniques and represent unimportant causes of amputation today.

Should an elderly patient have an artificial leg? An affirmative answer requires consideration of many factors. In the young, the primary decision for prosthetic replacement usually rests on local conditions; in the aged there are systemic considerations of equal importance. Every elderly amputee should have a complete physical examination as part of the evaluation to decide the feasibility of a prosthesis. However, cardiorespiratory disease, allowing nothing but mild physical activity (heart classification 3C or worse), will preclude sustained ambulation but not necessarily prescription of a functional prosthesis. Heart disease per se is not a contraindication.

---

*Associate Attending in Charge, Physical Medicine & Rehabilitation, Beekman-Downtown Hospital; Assistant Clinical Professor of Physical Medicine & Rehabilitation, New York University-Bellevue Medical Center; Director of Physical Medicine & Rehabilitation, Department of Hospitals, New York, N. Y.

**Director of Physical Medicine & Rehabilitation Department, Michael Reese Hospital, Chicago, Ill.

Some difficulties arising in the anatomic site are no different from those encountered in amputations in all age groups. Thus pain due to adherent scars, neuromata, exostosis, or infection must be prevented or relieved. Poor anatomic conditions due to inadequate shrinkage, dog ears, and the like, complicate proper fitting. Breakdown of skin from local allergies to wool or lacquer may be a vexatious problem. Other local difficulties are peculiar to the aged. Vascular deficiency must be considered when poor skin condition persists and then reamputation may be in order. Contractures of proximal joints represent road blocks to prosthetic substitution, since the elderly are more prone to this preventable complication, which occurs all too frequently because of poor nursing care. Reduction in strength by disuse is another common event in the aged, not only in the stump but generally. A reconditioning program is more helpful for this deficiency than for contractures.

OL.
4
956

Vascular or other changes of the central nervous system may result in a decreased sense of balance and poor muscular coordination. The most common complication of this type is hemiplegia. As a general rule, the hemiplegic patient is less disabled when amputation and paralysis occur on the same side. Experience has shown that this combination does not necessarily militate against successful walking. A functional test for evaluating motor capacity of such patients is afforded by their ability in learning the swing-through-tripod crutch gait. Intellectual loss from cerebral disease added to defective motor coordination may make adequate training impossible.

Severe loss of vision or even complete blindness is not considered an absolute contraindication for a prosthetic prescription. However, since touch and position sense of a living limb cannot be supplied by a mechanical device, training of blind persons is extremely difficult.

Most vascular diseases of the lower limbs are symmetrical. Therefore, the circulation of the remaining limb may be expected to be deficient. This by itself is no contraindication for limb fitting. On the contrary, prosthetic training should be encouraged against the day of possible loss of both limbs, inasmuch as an elderly double amputee has a much greater chance to master ambulation if he has already gained facility as a unilateral amputee. Double amputees with no previous experience in use of artificial limbs are usually doomed to failure.

Psychologic considerations are of great importance. Although patients almost always express the desire for replacement of a lost leg, this attitude may remain merely on the verbal level; actually they withdraw from any realistic attempt to master the use of a prosthesis. While these psychologic implications apply to any amputee, age may predispose to poor motivation because of loss of adaptive resiliency. Defeatism and apathy tend to be more pronounced. Sometimes rejection of a limb may be more subtle: while the patient makes a surface show of complying with social attitudes by receiving a limb, he finds an endless round of faults against wearing it. One must be alert against this type of unconscious invalidism and differentiate it from real deficiencies of stump and socket. Such a patient is suffering primarily from a psychologic disorder rooted in the inability to reconcile the loss of a member. Closely related to this category is the phenomenon of "phantom limb," rare but most disabling.

It is not implied, however, that psychologic manifestations are necessarily grave. It is expected that a period of grief will supervene following the loss of a limb as it does following the loss of a member of the family. This is natural and without it adjustment cannot ensue. Consequent to amputation, therefore, depression and other emotional manifestations of psychic trauma should not interfere with the prosthetic procedure. In fact, resumption of the biped position is often an important component of psychotherapy.

The choice of site of amputation is a crucial decision. Preservation of the anatomic knee joint facilitates rehabilitation even of the most athletic individual. In the group under discussion presence of the knee joint is not only a question of ease of ambulation but frequently determines the feasibility of an artificial limb. It is hoped that new drugs will enable surgeons to perform operation at lower sites more frequently.

The best surgical and rehabilitation techniques still leave a hard core of patients, mostly above-knee double amputees, who become wheel-chair patients. It is often disregarded that amputees need special wheel chairs. With the weight of lower extremities eliminated, the center of gravity of a sitting patient is shifted backward. This can be corrected either by placing weights in front of a standard chair or preferably by obtaining a special chair with a long base produced by moving rear wheels back. Unless these precautions are carried out, there is a danger of the chair tipping backward, resulting possibly in a head injury.

An amputee chair should be equipped with brakes, upholstered arm rests, and at least a 2-inch foam rubber cushion. These few minor conveniences are of the greatest importance.

The suction socket has been an important development in prosthetics. Elimination of a pelvic band is a convenience to the elderly, especially as obesity and hernias are more prevalent. The deciding factor in prescribing a suction socket for this group of patients is the strength of upper extremities necessary for pulling the stump sock through. An absolute contraindication is muscle weakness and resulting loss of muscle mass of the affected limb. Age by itself is not a primary contraindication.

Regardless of the type of prosthesis, the eventual goal of rehabilitation is a patient who is able to walk without the aid of crutches or cane. This may be applied fairly rigidly in young individuals; in the aged, compromise is often unavoidable. Ambulation on a prosthesis, even with the aid of two crutches, is definitely an improvement over wheel-chair existence. Similar compromise extends to the problem of knee stability. Many an elderly amputee will be discharged home with the knee lock in place. At a younger age the advantage of knee mechanism can be utilized in almost all patients. In the presence of impaired coordination, mobility has to be sacrificed for stability.

Although advances have been made in medical and prosthetic management, the economic factor is important. The income of the older age group is below the national average. Medical expenses previous to amputation have often used up the available financial resources. Unless the cost of a prosthesis and rehabilitation can be provided, the patient may be forced to leave his home

surroundings and to become a charge of the community. The year in, year out, cost of custodial care will be multiplied many times compared to the single outlay for rehabilitation.

The Office of Vocational Rehabilitation provides many prostheses for employable individuals. However, the aid of this agency can be utilized only infrequently. The vocational possibilities of the elderly are limited in the present labor market. In New York City, the rehabilitation-minded Department of Hospitals and Welfare have been most cooperative in providing funds for prostheses and training of amputees. It is to be hoped that agencies in other municipalities will follow this example. All major expenses of public monies should be properly supervised. Such a safeguard is exercised in New York City by limiting specialists in Orthopedic Surgery and Physical Medicine and Rehabilitation to writing prosthetic prescriptions for municipal agencies.

Some amputees are unable to utilize functional prosthetic devices, and are wheel-chair bound. Consideration should be given to these individuals in the form of an inexpensive cosmetic limb, since the psychologic value of such a device cannot be underestimated. To the best of our knowledge such an inexpensive cosmetic replacement is not available.

OL.
4
956

SUMMARY

All elderly amputees should have a complete medical examination prior to prosthetic fitting, as the general state of health commonly determines the success in using a prosthesis. Physiologic, not chronologic, age is the decisive factor.

Complaints rooted in unfavorable personality characteristics may be mistaken for pain due to physical causes. In such cases the best medical and prosthetic skills alone will be of no avail unless adequate psychiatric help is provided.

There are few absolute contraindications to prescription of a prosthesis: these are severe cardiovascular disease, inadequate motor coordination, and senile mental changes.

The economic problems of the aged deserve as much attention as the medical. Community facilities should be available for the purchase of prosthetic devices and training in their use.

Review

SOME DEVELOPMENTS IN CANCER RESEARCH

G. Burroughs Mider, M.D.

Bethesda, Md.

From the National Cancer Institute, United States Public Health Service, Department of Health, Education, and Welfare

VOL.
4
1956

(Received for publication April 23, 1956.)

AN EMINENT student of cancer, asked to comment on a large international meeting, replied, "The program consisted of papers on clinical cancer and papers on cancer research." This unguarded response can be interpreted in various ways, but all of them leave the unmistakable impression of a distinct dichotomy between clinical and laboratory investigations of neoplastic diseases. The practicing physician measures progress in cancer research by improvements in his ability to manage cancer as it occurs in his practice. The public health official is impressed by significant changes in vital statistics of cancer. The laboratory scientist is convinced that painstaking accumulation of data on normal and abnormal growth will eventually provide practical means of cancer control, which, in its final analysis, depends on prevention or effective therapy of cancer in man. If one were to carry the analysis further by isolating the various orthodox disciplines of medicine and science represented in each group, a distinct inclination for individual experience to circumscribe one's outlook on cancer research would become obvious. These different attitudes merely reflect diverse philosophies of the various disciplines concerned with a common problem. Each is looking at different faces of the same enormously complex polyhedron. Specialization creates problems. It also solves them.

The past twenty years have seen a greatly augmented effort to achieve control of cancer through clinical and laboratory research. New talent has been recruited, new facilities provided. Unstinting support has flowed from public and private philanthropy. The whole has created a valuable national resource, but not a national or international program in the sense that the common goal is being approached through central coordination of individual or group efforts. The control of cancer requires pioneering beyond the frontiers of human knowledge, as does any proper subject of research. This is in contradistinction to the engineering type of approach that assumes practicality when the application of known pertinent, basic principles to a particular problem can achieve the desired results. The total experience of all engaged in investigations of

cancer in the clinic and in the laboratory gives direction to the efforts. Emphasis shifts as scientists and clinicians explore promising new leads that may be exploited for the benefit of the cancer victim. Fund-granting agencies are quick to provide support for concentrated work which evaluates some clue of potentially practical moment.

An excellent example of such flexibility relates to the earlier experience in treating lymphomas and allied diseases with cortisone and adrenocorticotrophic hormone of the hypophysis. Both of these materials were difficult to obtain at the time when dramatic therapeutic responses of certain neoplastic diseases to their administration were first reported.[116] The Public Health Service obtained a small quantity of both cortisone and ACTH through the cooperation of the pharmaceutical industry. A special committee parceled out the materials to competent investigators for the express purpose of evaluating the therapeutic usefulness of the agents in a variety of neoplastic states. It took little more than a year to show that each hormone was useful in the management of certain lymphomatous and related diseases. Neither offered hope of cure nor were they in any sense panaceas for the generality of cancers. The final word on cortisone and ACTH in the treatment of malignant neoplasms remains to be written, but critical analysis of the general problem was approached effectively.

THE GENESIS OF CANCER

Practical prevention of clinical cancers requires better knowledge of their causation than exists today, yet a pitifully small fraction of cancer research, in the United States at least, is primarily concerned with this extremely important aspect.[138] The basic facts on the etiology of cancer have changed little since they were summarized by Morton in 1937.[109] Carcinogenic potency is not confined to any one class of agent or to any particular type of chemical compound. The ability of a carcinogenic agent to produce cancer is, furthermore, conditioned by certain heritable traits in laboratory animals. Genetic influences probably play no greater etiologic role in neoplastic diseases than they do in many other types of illness,[39] but reproducibility of response is achieved best in the study of experimental cancer through careful control of genetic constitution of the experimental subjects. Resistance and susceptibility to cancer, as determined by heritable traits, are relative. An overwhelming stimulus may completely mask resistance, or extreme susceptibility may produce as many neoplasms of a given type without known stimulation as with it, as long as the frequency of neoplasm is the only point of reference. It takes a long time to produce cancer with most carcinogenic agents—long in terms of the life span of the species under investigation—and the variation in length of time between exposure to the cancer-producing agent and the definitive response, the latent interval, is often a reasonably good index of inherent susceptibility or resistance, on the one hand, or relative strength of an inciting agent, on the other. Individual members of an inbred strain of mice are almost as uniform as are identical twins. Their incidence of specific anatomic types of neoplasms are predictable within comparatively narrow limits. Transplantation of pieces of organs varyingly susceptible to spontaneous or induced tumors, to hosts of different inherent

susceptibility, reveals that the probability of a neoplasm arising within the transplanted tissue is determined by the constitution of the donor.[72,119,128] The new environment has little influence on the neoplastic potentialities of the graft. Probably the principal action of hereditary influences in tumor formation is localized in one or more organs or tissues of the body rather than representing an over-all influence on the total organism. Such reasoning is in keeping with the spontaneous occurrence of cancers of specific organs among inbred strains of mice. Tumors of the lung, mammary gland, and hematopoietic tissues do not appear with equal frequency in any one strain. Genetic influences determine in large measure both the individuality and the reaction of the individual to its environment; hence, definition of carcinogenic environmental factors may be the most practical means of approaching cancer prevention. Correlations between occurrence of certain types of cancer and morphologic, physiologic, or biochemical events, as determined by controlled breeding of laboratory animals,[71] may disclose situations that can be taken to the clinic for affirmation or denial in human experience. Physiologic, or biochemical, genetics is a healthy infant whose continued growth promises much for many fields of biology and medicine.

Recognition of groups particularly prone to development of one or another specific anatomic type of cancer among the population is extremely important to a better understanding of the causation of neoplastic diseases. Our current knowledge results from the efforts of individual physicians who happened to be in the right places at the right times, and the careful analysis of their experience. An alert, interested clinician will suspect that cause and effect may be operating if his practice contains a preponderance of patients with a common occupation or habit who develop the same disease. The method is limited, however, to chance observations, and the urgency of the situation requires more vigorous pursuit of the information we need.

Epidemiologic studies continue to be useful in control of infectious diseases. The methodology already available is being adapted to the systematic study of cancers as they occur in racial, geographic, occupational, and other groups. New techniques are being developed with the objective of defining inordinate risks to particular cancers among the populace, which, it is hoped, will lead to definition of etiologic agents that may be excluded from the environment.

The total effort of all interested parties has produced, over the years, only six clearly identifiable agents that are known to cause cancer in man: (1) x-rays and some of their constituent parts (whether generated by machines or by artificially induced or naturally occurring radioactive substances makes little difference if experimental data are correct), (2) ultraviolet irradiations, (3) arsenic (as potassium arsenite, usually after prolonged medication), (4) beta-naphthylamine, (5) benzidine, (6) 4-Aminodiphenyl.[102]

The same general principles of carcinogenesis seem to apply in the clinic and the laboratory, but the specific anatomic type of cancer produced in experimental animals may differ from the type of neoplasm obtained from accidental exposure of man to the same compound. Beta-naphthylamine, for example, is a potent carcinogen for the bladder in man and dog. It produces hepatic carcinomas in mice. Clayson, Bonser, and Jull[23] present evidence that beta-naphthylamine is metabolized differently by various species, and the relative

VOL. 4 1956

proportion of 2-amino, 1-naphthol conjugates formed in metabolic degradation by a species can be correlated with its liability of developing neoplasms of the urinary bladder. Introduction of paraffin pellets containing beta-naphthylamine into the bladders of mice produced no cancers, whereas incorporation of 2-amino, 1-naphthol-HCl into similar pellets evoked cancers in the urinary bladder. Some tumors resulted from exposure to paraffin alone, but the much longer latent interval of tumor production by paraffin suggests that the naphthol compound is a considerably stronger carcinogen.[13] Correlation between human and animal experience, though generally excellent, is not absolute. No arsenical preparation has yet been shown to induce cancer in experimental animals.

A much larger number of substances, essentially mixtures, has been incriminated in the production of cancer in man.[77] Soots,[69] shale oils,[69] paraffins,[117] undefined materials associated with the production of nickel,[103] chromates,[98] and isopropanol[150] have increased the hazard of developing one or more types of cancer among persons exposed. This list is not complete, but interest in the others is far overshadowed by a suspected etiologic relationship between certain environmental factors and the increasing frequency of pulmonary cancer. Surely carcinoma of the lungs and bronchi is listed more often as a cause of death than it was twenty-five years ago. The burden of evidence indicates a real increase in the incidence of primary pulmonary neoplasms in the United States[37,38] and in Western Europe.[24,78,143] The environmental basis for this development seems strong because radical changes in the biologic characteristics of the diverse peoples involved could hardly be accomplished during a single life span.[76]

Increase in pulmonary cancer has been much more pronounced among males than females and is more evident among urban than rural residents. The magnitude of the change and its widespread geographic distribution suggest that the occupational or industrial practices known to be associated with an increased risk of carcinoma of the lungs and bronchi cannot be primarily responsible. The carcinogenic agent or agents must be widespread indeed. The atmosphere is contaminated by the combustion products of various fuels and by the effluvia of industrial processes. Soots collected from the air of cities have long been known to possess carcinogenic potency for mice.[94,95] Benz[*a*]pyrene, formerly 3, 4-benzpyrene, a potent carcinogenic chemical for experimental animals, can be isolated from the atmosphere. It is a pyrolysis product of many carbonaceous materials. Strong circumstantial evidence incriminates this compound as the principal causative factor in the development of chimney sweep's cancer, first described by Pott,[118] but absolute proof is lacking. Waller[146] has identified benz[*a*]pyrene in the air of English cities and has studied the effects of atmospheric conditions on its concentration. Attempts to isolate the chemical from human lungs have been unsuccessful.[10] Experience indicates that mixtures containing this material often evoke cancer in experimental systems but does not exclude the possibility that other equally important and powerful carcinogens may be present. Additional carcinogens chemically unrelated to the polycyclic hydrocarbons have been isolated from the atmosphere by Kotin, Falk, Mader, and Thomas.[80] Some of them are present in the particulate phase of gasoline-engine exhausts.[81] There can be little doubt that common human experience includes unintentional exposure to a number of agents that can evoke some cancers in some animals.

An association between tobacco smoking and pulmonary cancer was first suggested by Müller in Germany.[110] Extensive retrospective studies of smoking habits of patients with carcinomas of the lungs and bronchi[36,96,125,127,154] have uniformly agreed that a larger proportion of these people have smoked "excessive" numbers of cigarettes over a long period than have persons without pulmonary cancer who comprised the control groups. The duration of significant exposure varies somewhat in the different investigations, as do the proportions of heavy cigarette smokers in the lung cancer groups. Statistically significant associations between other forms of tobacco consumption and other types of neoplastic disease have been reported.[125,127] Sadowsky, Gilliam, and Cornfield[125] converted their findings into estimates of prevalence rates of pulmonary cancer in smokers and nonsmokers. The rate in smokers was 5 times that in nonsmokers, but others have found as much as a 14-fold increment. Regardless of the conservatism of the printed reports, the belief that an increased risk of the development of pulmonary cancer has been demonstrated among cigarette smokers is widespread. Prospective studies necessary to decide the point have not yet been completed; preliminary findings are inconclusive.[67] Interpretation of the statistically significant association between excessive cigarette smoking and pulmonary cancer as adequate evidence of a cause and effect relationship of major import in the increase in cancer of the lung seems unjustified without other substantiating data. Hammond[66] argues that the latter are available from laboratory studies.

Wynder, Graham, and Croninger[155] have summarized results of investigations on the carcinogenicity of cigarette tars formed by condensation of the smoke or extracts prepared from them. While tumor production followed application of these materials to the skin of mice for at least one year, most of the lesions recorded in the literature have been papillomas rather than cancers. Wynder and his associates[155] observed 44 carcinomas of the skin among the 81 mice they painted thrice weekly with an acetone solution of cigarette tar condensate for a maximum of two years. The first cancer appeared during the forty-second week and the average time of appearance was 71 weeks. One is tempted to compare the over-all carcinogenicity of this type of stimulus with other materials capable of evoking cancer of the skin in mice. Investigators are necessarily dealing with complex mixtures. Coal tars may be comparable but their carcinogenicity varies widely, depending on such factors as the source of the coal and the temperature at which it is burned.[149,153] The only potent carcinogen identified thus far in coal tars is benz[*a*]pyrene, but the carcinogenic potency of the mixture need not depend solely on the concentration of benzpyrene, since other substances lacking the ability to produce cancer may modify the response to a powerful carcinogenic agent.[8,30,115,126] The nature of the cancer-producing factor in condensates of cigarette smoke is uncertain. Aromatic hydrocarbons, including benz[*a*]pyrene, have been identified in it by Cooper and his colleagues.[27,28]

Few pulmonary cancers at all comparable to human types of bronchogenic carcinoma have been produced in laboratory animals. Mice do get a peculiar type of neoplasm of the lung which sometimes metastasizes, even though some pathologists refer to it as "adenoma." The ability of some agents to increase the frequency and decrease the latent period of formation of the pulmonary

VOL 4 195

tumors, and of certain other neoplasms to which the laboratory mouse is spontaneously susceptible, is often an extremely sensitive test of carcinogenic potency. Lorenz and his co-workers[97] exposed strain A mice to atmospheres heavily contaminated with tobacco smoke (no paper included) for 250 days and observed no increased incidence of pulmonary neoplasms. Essenberg[42] found significantly more pulmonary tumors in strain A mice that had been exposed to cigarette smoke for at least a year. These experiments should not be dismissed lightly as simply contradictory results. The duration of stimulus and the character of the smoke may have been important in determining the findings. The techniques used may be criticized as being more related to experience in a smoke-filled room than to actual smoking as practiced by man.

OL.
4
956

The evidence incriminating atmospheric pollution either by combustion products of fuel and industrial effluvia or by cigarette smoke is essentially similar. Statistical associations between pulmonary cancer and cigarette smoking or urbanicity are established. Soot or a chemical compound identified in soot can cause cancer in laboratory rodents; so can condensates of some tobacco tars, and one known potent carcinogen can be found in tar from cigarette smoke. The problem is to assess the relative importance of each of these factors in the increased incidence of pulmonary cancer in man, with the full realization that either or both of them may ultimately prove to be red herrings. Reduction, if not complete abolition, of atmospheric pollution is probably within the competence of engineers and should be accomplished as a general public health measure irrespective of its possible relationship to cancer. The writer, a heavy cigarette smoker himself, finds no evidence that smoking materially improves physical well-being, no matter how much satisfaction the habit may afford. Both experimental and clinical experience indicate that carcinogenic agents may act at sites distant from their portal of entry to the body. The lung is an important excretory organ and the possible production of pulmonary cancers by agents entering the human body through extrarespiratory channels cannot be disregarded.[76]

An excellent discussion by Oberling and Guérin[111] of the role of viruses in the production of cancer begins: "It is now more than 50 years since Borrel founded the hypothesis that cancer has a viral origin on the frail foundation of reasoning by analogy. This hypothesis has had a curious destiny. It was so contrary to everything that was known—or was thought to be known—that from the beginning everyone rallied against it. Nevertheless the facts in its favor have continued to increase slowly but surely, until at the present time the virus hypothesis has become a crucial point in discussions on the etiology of cancer." One may quarrel with the authors' statement that the virus theory "offers the best explanation for malignancy," but steady progress in producing mammalian neoplasms by filtrable agents obtained from cancers suggests the desirability of intensified research in this field.

The cancerous responses to viral agents had generally been more rapid than those to chemical carcinogens until Bittner[9] established the presence of a factor in maternal milk that influenced the ultimate incidence of mammary carcinomas in mice some months later. The milk agent has many of the attributes usually associated with a virus. Its action is conditioned by various

factors. An extensive review of this interesting subject has recently been published by Dmochowski.[35] Search for a comparable agent in the genesis of human mammary cancer has been fruitless to date and, indeed, mice do develop carcinomas of the breast in the absence of demonstrable quantities of the milk factor.[70] The mammary tumor inciter is widely distributed among the tissues of mice even though it acts solely on the mammary gland, can be transmitted to the female by spermatozoa, and can be recovered from families of mice in which mammary carcinoma occurs rarely.

Isolation of a filtrable agent causing lymphomatosis in chickens by Burmester, Prickett, and Belding[21] provides an additional virus that determines eventual incidence of another neoplastic disease. Lymphomatosis has assumed increasing economic importance to the poultry industry, especially in the production of pure breeds of fowls. Identification of the causative agent has shown conclusively that it can be transmitted through the egg even though its results may not be manifest until months later.[29] Waters and Bywaters[148] have demonstrated infection by contact with sick chickens, but only animals less than 40 days old are susceptible.

The distinct tendency for specific anatomic types of cancer to become concentrated in family lines by inbreeding has provided extremely useful strains of mice in which the occurrence of neoplastic diseases is highly predictable both as to kind of cancer and the age at which it occurs. One may argue that inbreeding so concentrated the genic material as to allow the ordinary vicissitudes of earthly existence to determine the reactions of the subject, including diseases to which it was heir. Gross,[58] however, took a different view by assuming that familial occurrence of malignant tumors could be explained by transmission from one generation to another of pathogenic agents capable of producing tumors. The hypothetical agents necessarily exist in inactive form until activated by some unexplained mechanism during later life.

Gross[59] tested his hypothesis by injecting into strain C3H mice (with low spontaneous frequency of leukemia) cell-free extracts prepared from leukemias arising in AK mice (in which the risk of leukemia is extremely high). This procedure caused leukemias to appear among the recipients when 8 to 11 months old, but only if the injections were made within a few hours of birth. The procedure has been repeated many times by its originator with consistent results,[62,63] yet attempts to reproduce his findings have not met the same degree of success.[84,140] Stewart[141] finds that injection of cell-free material from leukemias into first filial generation hybrids derived from strains of mice with high and low risk to the development of spontaneous leukemia increases the frequency of the disease and causes it to occur at an earlier age. To some, this may seem essential confirmation of the findings reported by Gross, even though Stewart's[140] experiments, designed to reproduce his original work, failed to yield similar results.

During the course of their respective investigations, Gross,[60,64] Law and his associate,[84,86] and Stewart[140] have each noted the occurrence of peculiar neoplasms of the salivary glands in mice receiving cell-free extracts or filtrates of leukemic tissues. These cancers had not been described previously, and are unequivocally associated with the experimental procedure. Thus the same

procedure has yielded at least one result on which all are agreed. Perhaps an explanation for the unconfirmed observations of Gross is afforded by the nature of the test animal used. Inbred strains of mice have been known to develop sublines of different genetic constitution as the result of spontaneous mutation or some other unrecognized type of biologic cataclysm, even though their general characteristics have not changed. It would appear that Gross has used C3H mice obtained originally from Bittner and from Andervont. The two lines have been maintained separately in his laboratory and the great majority of leukemias have occurred in the Bittner stock.[65] Law and Stewart have worked exclusively with National Cancer Institute lines of C3H mice closely related to Andervont's strain.

Most recent information implicates cell-free extracts with the genesis of neoplasms of the adrenal medulla and subcutaneous tissues.[64,86,140] The same procedure used to produce tumors of the parotid yielded just as many adrenal medullary tumors in an experiment reported by Stewart.[140]

OL.
4
956

These experiments imply that latent infection with carcinogenic viruses may be widespread. How specific may be the localization of viral action remains to be determined despite the apparent ability of Gross[61,62] to separate agents responsible for leukemia and salivary gland tumors, respectively. The role of nonviral carcinogens may require some re-evaluation. Rous and Kidd[79,124] demonstrated the ability of coal tar extracts to localize the action of the highly cytotropic Shope papilloma virus and to accelerate the transition of benign to malignant in the papilloma-carcinoma sequence. It does not seem necessary at this time to postulate a viral etiology for all malignant neoplasms, nor do the various carcinogenic viruses have particularly constant attributes other than their ability to evoke neoplastic disease in appropriate subjects.

Evidence for viral causation of a specific type of tumor rests almost exclusively on ability to produce that kind of lesion with a cell-free preparation from normal or neoplastic tissue. Positive evidence is unambiguous. Negative evidence may be inconclusive. Carcinomas resulting from injection of the Shope papilloma virus contain no directly demonstrable carcinogenic agent.[131] The virus tumors of fowls may pass through periods when transmission by filtrate becomes difficult or impossible.[111] Bryan, Calnan, and Maloney[17] have applied the methods of quantitative biology to a study of dose-response relationships of the Rous sarcoma virus. He finds very little or no agent in active form when the dose of virus produces neoplasms in less than 50 per cent of the chicks injected. The species specificity of those carcinogenic viruses now identified, though not absolute, is sufficiently restrictive to pose serious difficulties in approaching the study of viral etiology of cancer in man—for which no evidence exists today.

No matter how many agents may be capable of inducing the cancerous state, the immediate cause of cancer must be the production of a self-propagating, heritable abnormality within one or more cells. The prolonged period required to effect this change by most carcinogenic agents supports the concept of a stepwise process and, indeed, some investigators recognize periods of initiation, promotion, and progression[7] in the development of full-fledged cancers in certain experimental systems. The final transition to the neoplastic state may well

be a rather sudden event even though its overt manifestations require other changes either in the altered cells themselves, or in some host factor which promotes the chain reaction. Argument regarding the nature of the basic change has included consideration of some fundamental difference in nuclear constituents since many heritable traits are known to be a function of genic composition among sex cells. Genetics is based on the ability to demonstrate differences among phenotypes (depending on alteration in the germ plasm) which are, by definition, mutations. The somatic mutation theory of cancer, then, seems to be an extrapolation of this well-documented phenomenon to explain self-perpetuating changes that occur in somatic cells. Chromosomal and other nuclear abnormalities occurring commonly among cancerous tissues and an allegedly close correlation between carcinogenic and mutagenic properties of chemical or physical agents are proffered as arguments supporting this hypothesis, yet review of the evidence by Burdette[20] fails to support the latter hypothesis. Many of the bizarre changes in nuclear appearance seem to be lethal, and equal degrees of cellular atypism are found in some rapidly proliferating cells such as the developing trophoblast whose growth is self-limited. Genic material does reside in the nucleus, but its influence is more widespread. Beadle and Tatum[4] and other geneticists have shown distinct relationships between specific genes and specific chemical reactions known to require partipation of cytoplasmic constituents. Among neurospora, one gene may govern the activity of approximately one enzyme but single genes may affect more than one enzyme, and several genes may affect the same enzyme.[12] Extensive studies of neurospora and *E. coli* detected principally genes of unifunctional type.[74] Concepts of unipotent and multipotent genes are not mutually exclusive, as pointed out by Bonner.[11,12]

The precise nature of an individual gene is unknown, but its attributes suggest that it may be represented as a factor X which permits a chemical reaction to proceed, perhaps by controlling the availability of one or more of the reacting constituents. If the supply of cytoplasmic compound *a* is under genic control and compound *b* is supplied from the extracellular milieu, the presence of both in adequate quantity may lead to the production of compound *c* necessary to proper cellular function. Absence of factor X, *a*, or *b*, for any reason, will abolish the formation of compound *c*, whether it is due to some true genic mutation, failure of synthesis of a reactant, alteration of cellular permeability, or of some other mechanism by which the extracellular compound gains access to the intracellular environment.

Methods are not yet available to permit direct approach to the genic composition of mammalian cells, but mapping of chromosomes and their genes in relation to predictable risk of developing specific types of experimental cancer is proceeding. If true somatic mutation were the fundamental cause of cancer, one might anticipate a more or less random type of change resulting from the introduction of mutagenic agents. X-rays produce so many effects ascribable to chromosomal or genic reorientation that the responses may be considered randomized to some degree, yet dose-response relationships defined with respect to production of certain neoplastic diseases suggest a more systematic type of reaction. Bacterial geneticists have contributed importantly to the under-

standing of the production of heritable traits. The mutagenic agents employed by some of them include such simple compounds as manganous chloride which lead to relatively predictable changes in the biochemical attributes of *E. coli* in succeeding generations.[33,34] Some of these effects are reversible.

Studies of carcinogenesis by the Millers reveal that N, N-dimethyl-*p*-phenylazoaniline (formerly p-dimethylaminoazobenzene), a powerful hepatic carcinogen,[105] and benz[*a*]pyrene applied to the skin[104] are so tightly bound to intracellular proteins in the liver and epidermis, respectively, that they may be released only by alkaline hydrolysis, yet the neoplasms produced by these agents contain no protein capable of binding the carcinogens. Observation of the sequential changes in capacity of the liver to bind "butter yellow" prior to the onset of cancer indicates competition between the ability of the cell to form one or more specific proteins and the ability of the chemical carcinogen to bind them. Eventually, viable cells may be produced which have completely lost those systems controlling normal growth and represent initial tumor cells. The deletion hypothesis of carcinogenesis remains to be proved but constitutes a major advance in cancer research since it seems susceptible to experimental attack, even though chemicals which have no carcinogenic potency may also combine with protein. Sorof and Cohen found one electrophoretic class of soluble proteins markedly reduced in extracts made from hepatic neoplasms induced by 4-dimethylaminoazobenzene, although electrophoretically similar proteins were generally present in comparable extracts of normal liver.[136] Further identification requires substantial advances in fundamental protein chemistry, but the carcinogenic dye apparently moves with the "h" proteins in the electrophoretic field.[137] Subtraction of important intracellular constituents during carcinogenesis may be accompanied by an increase in some other component, for the cells do live. Rhoads has drawn an analogy between original and mutant strains of bacteria, and normal and neoplastic (mutant) mammalian cells.[122] Realization of the heterogeneity of nucleic acids has permitted some correlation between certain heritable traits and nucleic acid composition. Avery, MacLeod, and McCarty[3] converted offspring of Type II avirulent unencapsulated pneumococci to encapsulated forms by exposing them to pure deoxyribonucleic acid (DNA) obtained from Type III pneumococci, and DNA's derived from other sources have effected different transformations in other bacteria.[41] The possibility of a change in DNA (which resides only in the nucleus), influencing the behavior of the total cell, cannot be dismissed, and further work may clarify the precise interrelationships among intracellular constituents. Certainly neither nucleus nor cytoplasm can exist indefinitely when isolated from each other.

THE CURABILITY OF CANCER

"Early cancer is curable." The crux of that popularly supported statement lies in one's concept of the term "early." Textbooks generally lead one to believe that prognosis for cure decreases with increasing size of the cancer. This is probably true in those cases where all other things are equal, but rarely are they equal. Medullary carcinomas of the breast are often bulky masses, yet five-year survival rates are comparatively good.[106] Certain circumscribed

neoplasms of the stomach remain localized longer than do some much smaller gastric cancers.[139] Surgeons and pathologists are thoroughly familiar with widespread carcinomatosis starting from primary lesions as small as 1 cm. in greatest dimension, and almost every clinician has encountered obvious metastatic cancer in patients whose primary neoplasm defied detection by the most searching techniques available. Size alone, though of practical import as one factor in determining prognosis, can be most misleading in an individual case. Yet it seems wise to stress the importance of size to the lay public in order that medical advice may be sought promptly whether the lesion be cancerous or otherwise.

The limiting factor in effective treatment by established methods of most carcinomas and many sarcomas depends on the anatomic extent of disease at the time the lesion is treated. Distant metastasis vitiates effective local treatment, hence "early" assumes a biologic as well as an anatomic connotation. Few cancers, either in man or experimental animals, are preceded by constantly predictable changes as revealed to the histopathologist, yet the evolution of cancer from the competent provoking stimulus to the overt effect takes time, often months in the laboratory, and probably years in man. Evidence for additional evolutionary change after the lesion becomes recognizable as cancerous continues to accumulate since Greene described the ability of the fully "autonomous" cancer[52] to grow in the anterior chamber of the eye[51] or the brain[53] of foreign species. Furth[46] has summarized the available evidence on the transition from a state in which some neoplasms require rather specific situations for their progressive growth in genetically compatible hosts to a phase when such requirements are lost altogether. The bulk of the evidence deals with cancers arising in primary or secondary sex organs, since hormonal balances have pronounced effects on these tissues. Comparable data on other clinically important types of cancer are not yet available.

On the experimental side, the general situation is well exemplified by the studies of Morris, Dalton, and Greene[108] of thyroid tumors arising in mice after prolonged ingestion of antithyroid compounds. The capacity of these neoplasms to concentrate radioactive iodine (I^{131}) varies widely. Even from their inception, some of them lack any appreciable power to concentrate this element, while others take it up in an amount comparable to that stored by the normal thyroid gland. The experimental tumor achieves something approaching immortality by its serial transplantation in genetically appropriate individuals and, under these circumstances, the thyroid neoplasms once capable of concentrating iodine generally lose that ability altogether. The number of passages required for the change is completely unpredictable. The structural characteristics of the tumor cells need show no significant change.[107] One may argue that the small fragment serving as the transplanted inoculum is not completely representative, but the frequency of the phenomenon militates against this otherwise sound criticism. This experimental situation may be quite analogous to the patient whose primary thyroid neoplasm concentrates radioiodine in significant quantity but in whom the metastatic lesions fail to exhibit that faculty. One gains the distinct impression that cancers of laboratory mice often evolve new characteristics much more slowly in terms of life span than do those that affect man.

Typical epithelial neoplasms of pulmonary origin may acquire features characteristic of sarcoma on serial transplantation. While acquisition of sarcomatous patterns by frank carcinomas is extremely rare in clinical experience, many cancers affecting man appear progressively more anaplastic in their metastases.

Another example of change in the state of cancer cells may be inferred from chemotherapeutic experience. Those few types of malignant neoplasm that respond favorably, even dramatically, to drug action usually recur at some later date when they are often no longer influenced by the agent that once had a definite therapeutic effect. The prostatic carcinoma that once seemed to disappear almost miraculously after castration and/or estrogen therapy usually returns, and at that time female sex hormone is no longer a useful drug. The responsive phase of the cancer of the prostate may represent neoplastic cells with certain biochemical qualities so closely allied to normal prostatic epithelium that they are susceptible to the same deleterious influences. The ultimate acquisition of drug resistance would indicate some drastic intracellular biochemical change whereby attributes of normal prostatic epithelium, in so far as they pertain to estrogenic effect, are lost. This is conjecture unsupported by objective data at present, and, indeed, comparable experience with acquisition of drug resistance in experimental leukemia offers alternative explanations.

Several investigators have observed the development of resistance in mouse leukemias to one or more drugs which originally prolonged the lives of the leukemic hosts significantly.[19,85] The frame of reference in laboratory studies—prolongation of life per se—is somewhat different from clinical and hematologic amelioration in the patient. Resistance is evinced by failure of the same drug which prolonged survival to do so in subsequent transplant generations of the treated tumor. Law[83] has shown that a mass of leukemic cells never before treated with any therapeutic agent will yield some completely resistant leukemic neoplasms when small fragments of the mass are distributed among a number of genetically appropriate mice. This result is exactly comparable to the observation that a colony of bacteria may contain forms resistant to antibiotic action, and the technique of transplantation of the leukemic cells is analogous to the subculturing of bacteria. A cancer represents a population of cells in varying stages of growth, development, and, perhaps, disintegration; hence one is not surprised to find cells of different heritable traits, especially after long-continued propagation beyond the life span of the original host. Doubtless, important contributions would be made by studying the reactions of spontaneous and induced leukemias to drugs in their original hosts or only a few transplant generations removed. At any rate, Law's experiments show that drugs, in killing off susceptible neoplastic cells, may permit resistant forms to propagate their race—a process of selection. These studies do not eliminate the possible contribution of the drug itself to the resistant state, a phenomenon also recognized in microbiologic research.

Goldin and his associates[48] have greatly prolonged the lives of leukemic mice by administering "anti-folic" compounds. Those subjects which succumb eventually to their disease do not appear to do so solely because the cells have become refractory to drug action. Transplants from their leukemias to mice

of the same inbred strain develop as they formerly did and exhibit the same growth retardation by the same dosage schedule of the same antileukemic compound. This seems best interpreted as evidence for some important role played by the host's metabolism in drug action. Chemotherapeutic research is clearly contributing to our basic knowledge of cancer.

The final example of important change within recognizably cancerous tissue relates to carcinoma of the uterine cervix. A lesion once described as precancerous hyperplasia of that organ is now called carcinoma in situ, intraepithelial epidermoid carcinoma, Bowen's disease, or carcinoma of the cervix, grade O. In any case, one refers to a condition in which cells in the stratified squamous epithelium of the cervical mucosa assume the cytologic characteristics usually associated with frank cervical cancer but are strictly limited to their normal confines.[44,75,120] The lesion may persist as such for years, as has long been known. It is difficult to diagnose by inspection or palpation, but cases are discovered through exfoliative cytologic examinations as originally devised by Papanicolaou, Traut, and Marchetti.[113] Some apparently authentic cases seem to revert to a normal pattern without any known cause, particularly when the lesion accompanies pregnancy.[54,68] Cervical carcinoma in situ may progress to invasive carcinoma which can metastasize widely. (Superficial cancers seemingly limited to mucosae in other organs, notably the stomach, do possess the power to metastasize through vascular channels without demonstrating locally invasive qualities.) Arguing by analogy, then, one may assume that morphologically cancerous cells localized to the cervical mucosa acquire the power to invade and metastasize.

The remarkable ability of neoplastic cells to change their behavior with or without concomitant structural change implies a spectrum of biologic types of cancer within the same anatomic, even histopathologic classifications. This general concept was developed by Ewing[43] in his classic, *Neoplastic Diseases*, and extended by Broders[14] in the histologic grading of specific cancers. Correlation between structure and function by pathologists has greatly facilitated many fields of medical practice. The microscope remains an extremely valuable instrument, but biologic and biochemical attributes of cancer cells deserve closer attention as means of further refining diagnosis and prognosis in individual cases. Water content, or respiration and glycolysis in vitro are known to vary widely among both clinical and experimental cancers, but no recorded attempt to correlate such variables with biologic properties of individual cancers seems to exist. The task is not particularly easy because of sampling difficulties. Extraneous material is often intermingled with tumor cells, but the remarkable ingenuity conventionally shown by scientific investigators should be able to surmount such obstacles.

Surgery and ionizing radiations (generated by machines or naturally occurring or artificially produced radioactive isotopes) remain the only definitive means of treating cancers. The past several years have seen critical reappraisal of some procedures that have become widely accepted. The value of radical mastectomy as the treatment of choice in cancer of the breast has been challenged.[99-101,114] Following radical mastectomy the five-year survival rate among patients whose carcinomas are limited to the mammary gland far exceeds that of patients in whom the mammary cancer has spread to the axillary lymph nodes. A primary lesion strictly limited to the breast should be eradicated by simple

VOL 4 195

mastectomy, hence the radical operation in such cases is unnecessary. The survival rate in the presence of axillary metastases is not significantly better when primary treatment is radical mastectomy or simple masectomy plus x-ray and/or radium treatment of the axillary deposits. It is also argued that generally accepted, definitive surgical treatment for mammary cancer is ineffective. The statistical data on which these arguments are based seem sound if duration of life from diagnosis and treatment of cancer to death from that same cancer is, of itself, an adequate criterion of therapeutic result. The treatment under such circumstances certainly failed of its objective but may have afforded an appreciable degree of palliation, particularly if no recurrence appeared in or near the operative site and the patient's arm remained comfortably useful. The general subject has stimulated much argument.[2,31,112,130,145]

General trends in surgical therapeusis have been toward more extensive operations, particularly in cancers of the uterus, stomach, lung, head, and neck. Better control of infection and attention to metabolic aberrations have greatly increased the practical limits of surgical operations, and experience alone can evaluate the benefits that may accrue to the cancer patient. Many of the procedures now being investigated are based on the distribution of metastases obtained from necropsy studies.[1,40] The value of such useful reports might be appreciably enhanced if the distribution of metastases could be correlated with duration of disease, presence or absence of the primary neoplasm, and such other attributes of the cancerous state as lend themselves to analysis. An extensive study of sites of visceral metastases renders untenable the ancient thesis that the spleen is rarely involved in disseminated cancer. The modern data may be weighted with patients whose lives were prolonged by improvements in general medical care. On the other hand, a growing primary malignant neoplasm may conceivably restrict the ability of its metastatic deposits to grow. This, for example, might explain the common development of pulmonary metastases following careful amputation, under tourniquet, of an extremity for osteogenic sarcoma. No direct evidence bearing on some of these points is available, but a substantial increase in knowledge of metastatic mechanisms has accumulated from study of experimental systems.

Coman and his associates[25,26,159] found the number of metastases proportional to the number of viable tumor cell emboli delivered into the circulating blood. The site at which the cancerous emboli were introduced determined their distribution. These workers recognized that metastasis depended to a large extent on unrecognized factors within the primary tumor. The cancers used were transplantable lesions serially propagated in their host species for years and probably represent fully developed aggressive cancers since some of them grow in randomly bred members of their species. Sugarbaker,[144] studying the distribution of metastatic deposits produced by intravenous injection of cancer cells, described characteristic patterns associated with each of the individual neoplasms he used. Watanabe[147] studied metastasizability of a number of spontaneous mammary cancers arising in one mouse strain. He found a direct correlation between the capacity of these neoplasms to produce metastatic deposits in the lungs of mice on intravenous injection and their ability to grow in a foreign host but not with their morphologic characteristics. Metasta-

sizability may imply not so much the ability of cells to enter a vascular channel as their power to survive and proliferate in the blood stream or in their new site. Calcium is responsible to some degree for cohesiveness of cells in aggregates. Cancers are generally lacking in calcium and it is proposed that this, plus their active ameboid action demonstrated in vitro, contributes to invasiveness, hence access to vascular channels.[32] Interstitial pressure within a cellular mass has been proposed by Young and co-workers[156,157] as an additional important factor in dissemination of malignant neoplasms. Zeidman and Buss[158] demonstrated conclusively that cancerous cells may traverse the pulmonary capillaries and be filtered from extracellular fluid by lymph nodes.

Theoretically effective surgical procedures come to naught all too often because of local recurrence of cancer in or near the operative site. Contamination of the operative wound by neoplastic cells has long been suspected of playing a major role in these unfortunate sequelae, but Smith and Hilberg[135] have actually demonstrated the occurrence of unmistakably cancerous cells, even clumps of them, in washings of wounds created by dissection en bloc as a definitive procedure for clinical cancer. The extent to which wound seeding occurs deserves far greater exploration as a means of learning more about the biologic attributes of cancers in man, and the problem defined by the observations of Smith and Hilberg will undoubtedly lead to research on practical means of destroying the contaminants.

Despite limitations imposed on the surgeon by the biologic characteristics of cancer, he, with the radiologist, could deal more effectively with cancer if its victims could present themselves when their disease is limited to its site of origin. An all-inclusive diagnostic test for malignant neoplasms that may be expected to select cancerous subjects from an asymptomatic population has not been discovered.

The control of cancer by systemic administration of drugs or other agents is a highly desirable objective. Probably no other area of cancer research has received more attention during the past decade. Substantial information has been acquired. Drugs which physicians find useful in treating some patients with some malignant neoplasms have been added to the therapeutic armamentarium. Unfortunately the desirable effects of such agents, though often dramatic, are generally temporary so that progress to date has been toward better palliation rather than definitive treatment. Exploitation of the currently available knowledge is certainly indicated, but ample opportunity for developing other approaches to the problems of cancer chemotherapy exists.

An enormous effort to characterize the chemical attributes of cancer that are responsible for obvious differences in the behavior of malignant cells from their normal prototypes continues. A segment of the scientific community believes that therapeutic research will benefit most from this type of investigation. The chemical differences between normal and cancerous cells so far discovered have been quantitative rather than qualitative. Despite steadily increasing information on the biochemistry of cancer, brought together recently by Greenstein,[57] many facets of biochemical information remain virtually unexplored, not for lack of interest but simply because techniques must be devised to measure precisely ever-decreasing quantities of known entities and to discover

VOL 4 195

new reactions, even new compounds of biologic importance. Methods of enzyme chemistry still yield results reflecting capacity rather than actual physiologic state of activity. Detailed studies of isolated biochemical systems emphasize the organism's amazing ability to accomplish a single end by different routes. Complex interrelationships between many intracellular constituents remain to be defined. It seems just as unreasonable to criticize the chemist or morphologist for having failed to discover major differences between cancerous and normal cells to be exploited for therapeutic means as it is to emphasize failure to produce a truly effective drug that permanently eradicates cancer. Both attempts are of recent origin, both are making progress, and each offers much to the other.

Agents interfering with the oxidative utilization of carbohydrate by slices of neoplastic tissue have had little effect on parenteral administration to tumor-bearing animals or have proved extremely poisonous. Weinhouse[151] has shown the altered pattern of carbohydrate synthesis to be due to subtle differences rather than to absence of one or another important enzyme in the tricarboxylic acid cycle. Some pathways of glucose metabolism seem to be common to a wide variety of tissues and organs. No reason exists to believe that cancerous tissue would be any more sensitive to interruption of the major cycle than would noncancerous tissues. The problem may be similar to that encountered in development of chemotherapeutic agents for the control of microbial infection on which Krebs[82] has commented as follows:

> . . . In view of the similarities between energy-producing mechanisms of host and infective agents the prospects of finding selective agents are not very good. Such quantitative differences between host and parasite that exist are not likely to be a basis for differential chemotherapy.
>
> Energy production, of course, only represents one aspect of the metabolic activities of living cells, anabolic processes form another major part. Here again the pathways and mechanisms are generally the same in hosts and microorganisms, though there are differences in range: some substances may be synthesized only by the host or by the parasite. These qualitative differences may in some cases form a basis for chemotherapy, but what is more hopeful are the quantitative differences in the anabolic activities between host and parasite.
>
> The rate of energy production may be, say, tenfold in parasite, as compared to the host, and the growth rates of at least some organisms differ by a much higher factor from those of host cells. Moreover, many bacteria (at body temperature) must grow or die and therefore a major slowing down of growth is more serious to the bacterium than to the host. Efforts to interfere with anabolic processes may therefore be expected to be more fruitful than efforts to interfere with energy production.

The cancer cell is an alien. Although the degree of foreignness may be much less than that characteristic of a microbial invader, sound reasons exist for exploring the possibility of achieving equally satisfactory results from some of the approaches that have been used in developing antimicrobial agents. Rhoads[123] argues, "It has been hard for those who think only in terms of external invaders to conceive of the assumption, by the body's own cells, of new characteristics which render them functionally as foreign and aggressive in relation to their siblings and progenitors as are tubercle bacilli or the spirochetes of syphilis. This difficulty exists despite the fact that every morphologist has observed invasions of the tissues and of lymphatic and vascular channels, by

these neoplastic cells, has seen the growth of their secondary deposits in new locations, and has described the reaction of tissues to their presence. All these phenomena simulate in every respect local sepsis, bacteremia, and the tissue response to infection. The morphologic evidence, then, permits us, at least as an intellectual exercise and a guide to experiment, to regard the cancer cell as a foreign invador and to seek the cure of disseminated cancer by its selective chemical destruction, exactly as we cure the disease while localized by its selective mechanical removal."

A "rational" approach to chemotherapy is suggested by Rhoads,[122,123] who considers the heterogeneity of deoxyribosenucleic acids (DNA) the important point of departure. The chromosome controlling the reproductive equipment of the cell contains DNA, and it seems reasonable to assume that DNA may be different in normal and neoplastic tissues. "Since DNA of the chromosome is in dynamic equilibrium, at least prior to cell division, that of the neoplastic cell, as an abnormal one with abnormal dynamics, should be susceptible to antimetabolite effects to which the normal, related cytologic units are either not vulnerable or less vulnerable" In addition to the evidence for the important role of DNA in determining individuality as expressed above, it is also argued that ionizing radiations, and such drugs as nitrogen mustards and related compounds, antifolic agents and antinucleic acid precursors "direct their attack on the components of the chromosome."[123]

Experience with "antifolic" compounds and with 6-mercaptopurine in the treatment of acute leukemias has stimulated extensive study of the mechanisms by which these drugs act. Both seem to be concerned with altering the synthesis of nucleic acids or their precursors essential to cell growth and function. Folic acid (pteroylglutamic acid) is metabolized to citrovorum factor. The latter compound transfers carbon fragments as formate to the 2 and 8 positions of the purine ring, the balance of the purine skeleton being formed from carbon dioxide and glycine.[152] Studies with radioactive tracers have shown that antifolic compounds prevent the incorporation of C^{14}-labeled formate into the purine skeleton of the nucleic acids.[50,133] This is a net result, as Skipper[132] has emphasized, and does not reflect the specific chemical events affected by the antimetabolite. "It is obvious that any of a number of blocks in the energy-coupling processes, precursor-producing reactions, or catalytic events could provide for such slowing down of *de novo* nucleic acid synthesis." It is probably practical to define the precise role of antifolic agents in reduction of purine synthesis from small fragments using bacterial or other microbiologic systems and, indeed, in homogenates produced by disruption of mammalian cells. Direct extrapolation of such information to the intact animal whose neoplasm is severely damaged by administration of an antifolic drug is unwarranted without substantial evidence that the same events occur under in vivo conditions. After all, a dead cell can probably not incorporate any chemical into its protoplasm, hence failure to demonstrate any significant quantity of radioactively labeled constituent in the nucleic acid portion of cells from a treated neoplasm does not, of itself, constitute adequate evidence that the drug interfered primarily with nucleic acid synthesis. Citrovorum factor can give formate groups to other metabolically important compounds.[134]

The mammal can also utilize preformed purines to some extent in the biosynthesis of nucleic acids.[15] Free pyrimidines are not used in significant quantity. Systematic study of a large series of chemical relatives of the natural purine and pyrimidine bases of nucleic acids led to discovery of certain antagonists of nucleic acid synthesis in bacterial systems.[73] 6-mercaptopurine has been studied most extensively. Clarke and co-workers[22] found that experimental cancers treated with 6-mercaptopurine failed to grow on subsequent transplantation and also observed regression in the growth of some samples of sarcoma 180. Some clinical leukemias unaffected by antifolic compounds were influenced by systemic administration of 6-mercaptopurine.[18] Goldin and his group[49] subsequently demonstrated reversal of 6-mercaptopurine toxicity by compounds intermediate between purines and nucleic acids. Citrovorum factor, which abolishes toxicity of antifolic drugs, had no similar influence on 6-mercaptopurine toxicity. These observations suggest a mode of action for 6-mercaptopurine different from the action of antifolic compounds, but evidence for the precise mechanism of action by which 6-mercaptopurine produces its anticancer effect is incomplete.

VOL.
4
1956

Different species evidently utilize specific purine and pyrimidine bases somewhat differently in building polynucleotides, and data are now beginning to appear suggesting that some types of cells exhibit strong qualitative preferences for nucleic acid precursors.[5,6] An important role can certainly be assigned to nucleic acids in cellular metabolism. The precise nature of that role is still to be defined. The remarkable experiments reported by Fraenkel-Conrat and Williams[45] in which inactive nucleic acid and protein *sub*-units of tobacco mosaic virus were combined to produce active virus constitute a long step forward. Perhaps the most reasonable comment concerning the current emphasis on nucleic acid metabolism (in its broadest sense), as it relates to cancer, was published in 1953 by Skipper[132]:

> . . . The evidence that certain of the heterogeneous groups of agents which temporarily affect neoplastic growth also effect chemical events involved in nucleotide metabolism would seem to emphasize the necessity of attempting to obtain more basic knowledge in this field of biochemistry, if for nothing more than to lay to rest the growing assumption that here lies the core of the cancer problem.

The life and death of the cell, normal or cancerous, is concerned with many processes. Proteins must be built to provide enzymes for cellular function. Coenzymes and other molecules are essential to protoplasmic synthesis. While one hopes that the nucleic acid story may be written as a glowing monument to the control of cancer, this would not appear to be the time to concentrate all resources in that direction to the exclusion of other important avenues of approaching the arrest of cancerous growth.

Shear,[87-93,129] the father of systematic search for agents which damage cancers, has reported, with his colleagues, a large number of agents that produce unequivocal damage in mouse sarcoma 37. This experience should be regarded as a sort of litmus paper test for identifying chemical tools for further experimentation. No reason now exists to believe that each of these agents produces its effect in precisely the same way, though necrosis is a common denominator in each case. Experience with prolonged administration of α-peltatin clearly

indicates its ability to suppress growth of sarcoma 37 for a time, but escape from drug effect eventually supervenes.[56] Those few drugs developed by Shear and his associates[16,55,121] that have been introduced into the clinic have had interesting pharmacologic effects rather than therapeutically valuable ones.

The idea of developing a drug on the basis of known metabolic or biochemical facts appeals to the intellect, but only one of the many useful drugs available to physicians was developed strictly from that type of effort: British anti-Lewisite stemmed from Voegtlin's observation of the remarkable affinity of arsenic for sulfhydryl. The rest of them came from careful observations of individuals under a wide variety of circumstances, some so remote that they are traced back to folklore. Careful, critical observation is the basis for all research, and today's empiricism becomes tomorrow's science. The "rational" and "empiric" approaches to the chemotherapy of cancer are not yet mutually exclusive.

Chemotherapeutic research needs a better test object than now exists for predicting clinical experience. A large cooperative investigation of 27 chemical compounds, including some with known antitumor activity, some known to affect certain biologic systems other than tumors, and others of different particular interest, was conducted in 74 biologic systems.[47]

> It was hoped that such an examination would help to answer two very practical questions:
>
> 1. Does tumor growth inhibition by a compound reflect an effect on a property or function which is more characteristic of tumors than of other biological systems or is it an expression of an action on a property or function common to many diverse biological systems? If the latter is the case, can non-tumor systems be found which might be more useful, rapid, and economical than the experimental tumor systems?
> 2. To what extent does the study of inhibition of one experimental tumor have applicability to other experimental tumors?

The results failed to reveal a microbiologic, developmental or biochemical synthetic system which could replace an experimental tumor system as a screening tool for carcinostatic agents, and no single experimental tumor was found which could be expected to select all useful agents.

The failure of current screening techniques to predict clinical experience receives undue emphasis. Man is not a mouse nor are the neoplasms arising in the two species necessarily comparable in every respect. The bulky mammary cancers that arise so commonly in various inbred strains of mice are morphologically quite unlike the infiltrating duct carcinomas that constitute about 80 per cent of clinical cancers of the breast. The murine tumors remain localized, metastasize late, and seldom seriously discommode their hosts until ulceration or infection complicates the picture. Probably the leukemias, epidermoid carcinomas, and uterine cancers of mice are more closely akin to their human counterparts than are most of the other specific anatomic types. The leukemias, in particular, have proved most useful in forecasting clinical experience. The question as to whether spontaneous or transplantable neoplasms are preferable test objects fails to realize that each type of lesion may contribute its own kind of information. Methods now employed[142] have selected a few agents that are

clinically useful, and a compound seriously affecting the growth of an experimental cancer may also influence similarly some clinically malignant tumors. The most effective antibiotics might have been discarded prematurely had their attributes been examined on a series of microorganisms that resisted their action. It has been profitable to consider the many common denominators among different cancers and infectious diseases, but good reason exists now to focus attention on specific anatomic types of neoplasms and probably on even smaller subgroups within the larger categories.

A drug as potent in combating some forms of cancer as penicillin is in arresting the course of certain infections may ultimately be found. The search should continue. At the same time, some serious attention may well be given to exploiting the comparatively small effects of the anticancer agents already available. Correlations between chemical structure and biologic activity are already being sought. The fate and distribution of those compounds already useful deserve further study as a means of learning better how to use them. Perhaps some metabolic derivative of the drugs may be a more potent carcinostatic agent than the parent compound. Is the concentration of the agent in extracellular fluid important in determining its anticancer activity? These types of information have proved useful in other therapeutic research programs and may offer much to the progress of cancer chemotherapy.

OL.
4
956

The evaluation of potential anticancer compounds in the clinic is difficult at best. A large proportion of cancer victims have lesions that cannot be examined directly. Subjective improvement is often impressive; but the pay-off requires objective evidence that a neoplasm has been seriously influenced in a direction considered beneficial to its host. Increasing participation in cancer research by physicians skilled in internal medicine is encouraging, but the degree of that participation is still woefully small. *There is no one cancer problem.* There are many problems, including causation, attributes of the malignant state, therapeusis, epidemiology, and metabolic consequences of cancerous growth. This group of diseases, constituting one of the great frontiers of medicine and biology, provides unexcelled opportunities for exercising one's ingenuity, skill, and imagination. The ultimate prognosis is no worse than in many other diseases in which internists have an historical interest, and probably can be materially improved.

The very nature of some cancer problems requires their investigation by skills and tools available only in the experimental laboratory; other problems can be pursued only in the clinic, but many studies are prosecuted in both areas. The site of origin of a valuable research project is of no immediate concern, but progress in advancing knowledge depends upon crossing the street between laboratory and clinic when both can contribute to the common objective. It is a two-way street from which lead many alleys. It is present because no one building could possibly house all the different disciplines of medicine and science that are banded together to achieve the control of cancer any more than a single intellect can encompass the tremendous literature that reports progress in cancer research.

REFERENCES

1. Abrams, H. L., Spiro, R., and Goldstein, N.: Metastases in Carcinoma: Analysis of 1000 Autopsied Cases, Cancer **3**:74, 1950.
2. Ackerman, L. V.: An Evaluation of the Treatment of Cancer of the Breast at the University of Edinburgh (Scotland) Under the Direction of Dr. Robert McWhirter, Cancer **8**:883, 1955.
3. Avery, O. T., MacLeod, C. M., and McCarty, J. J.: Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types; Induction of Transformation by Desoxyribonucleic Acid Fraction Isolated From Pneumococcus Type III, J. Exper. Med. **79**:137, 1944.
4. Beadle, G. W., and Tatum, E. L.: Neurospora. II. Methods of Producing and Detecting Mutations Concerned With Nutritional Requirements, Am. J. Bot. **32**:678, 1945.
5. Bennett, L. L., Jr., Skipper, H. E., Stock, C. C., and Rhoads, C. P.: Searches for Exploitable Biochemical Differences Between Normal and Cancer Cells. I. Nucleic Acid Purine Metabolism in Animal Neoplasms, Cancer Res. **15**:485, 1955.
6. Bennett, L. L., Jr., Skipper, H. E., Toolan, H. W., and Rhoads, C. P.: Searches for Exploitable Biochemical Differences Between Normal and Cancer Cells. II. Nucleic Acid Purine Metabolism in Human Tumors, Cancer Res. **16**:262, 1956.
7. Berenblum, I.: Carcinogenesis and Tumor Pathogenesis, Adv. Cancer Res. **2**:129, 1954.
8. Berenblum, I., and Shubik, P.: The Role of Croton Oil Applications, Associated With a Single Painting of a Carcinogen, in Tumor Induction of the Mouse's Skin, Brit. J. Cancer **1**:379, 1947.
9. Bittner, J. J.: Some Possible Effects of Nursing on Mammary Gland Tumor Incidence in Mice, Science **84**:162, 1936.
10. Blacklock, J. W. S., Kennaway, E. L., Lewis, G. M., and Urquhart, M. E.: The Amount of Carbon in Human Lungs and Bronchial Glands, Brit. J. Cancer **8**:40, 1954.
11. Bonner, D.: Biochemical Mutations in Neurospora, Cold Springs Harbor Symp., Quant. Biol. **11**:14, 1946.
12. Bonner, D.: Gene-Enzyme Relationships in Neurospora, Cold Spring Harbor Symp., Quant. Biol. **16**:143, 1951.
13. Bonser, G. M., Clayson, D. B., Jull, J. W., and Pyrole, L. N.: The Carcinogenic Properties of 2-Amino-1-Naphthol Hydrochloride and Its Parent Amine 2-Naphthylamine, Brit. J. Cancer **6**:412, 1952.
14. Broders, A. C.: Carcinoma; Grading and Practical Application, Arch. Path. **2**:376, 1926.
15. Brown, G. B.: The Biosynthesis of Nucleic Acids as a Basis for an Approach to Chemotherapy, Ann. New York Acad. Sc. **60**:185, 1954.
16. Brues, A. M., and Shear, M. J.: Reactions of Four Patients With Advanced Malignant Tumors to Injection of a Polysaccharide From *Serratia marcescens* Culture Filtrate, J. Nat. Cancer Inst. **5**:195, 1944.
17. Bryan, W. R., Calnan, D., and Moloney, J. B.: Biological Studies on the Rous Sarcoma Virus. III. The Recovery of Virus From Experimental Tumors in Relation to Initiating Dose, J. Nat. Cancer Inst. **16**:317, 1955.
18. Burchenal, J. H., Ellison, R. R., Murphy, M. L., Karnofsky, D. A., Sykes, M. P., Tan, T. C., Mermann, A. C., Yuceoglu, M., Myers, W. P. L., Krakoff, I., and Alberstadt, N.: Clinical Studies on 6-Mercaptopurine, Ann. New York Acad. Sc. **60**:359, 1954.
19. Burchenal, J. H., Robinson, E., Johnston, S. F., and Kushida, M. N.: The Induction of Resistance to 4-Amino-N^{10} Methyl-Pteroylglutamic Acid in a Strain of Transmitted Mouse Leukemia, Science **111**:116, 1950.
20. Burdette, W. J.: The Significance of Mutation in Relation to the Origin of Tumors: A Review, Cancer Res. **15**:201, 1955.
21. Burmester, P. R., Prickett, C. O., and Belding, T. C.: A Filterable Agent Producing Lymphoid Tumors and Osteopetrosis in Chickens, Cancer Res. **6**:189, 1946.
22. Clarke, D. A., Philips, F. S., Sternberg, S. S., and Stock, C. C.: Effects of 6-Mercaptopurine and Analogs on Experimental Tumors, Ann. New York Acad. Sc. **60**:235, 1954.
23. Clayson, D. B., Bonser, G. M., and Jull, J. W.: An Experimental Inquiry Into the Cause of Industrial Bladder Cancer, Lancet **2**:286, 1951.
24. Clemmesen, J., Nielsen, A., and Jensen, E.: Mortality and Incidence of Cancer of the Lung in Denmark and Some Other Countries, in Clemmesen, J., Ed., Louvain, Council of International Organizations of Medical Sciences, 1953, p. 177.
25. Coman, D. R., Eisenberg, R. B., and McCutcheon, M.: Factors Affecting the Distribution of Tumor Metastases. Experiments with V_2 Carcinoma of Rabbits, Cancer Res. **9**:649, 1949.
26. Coman, D. R., deLong, R. P., and McCutcheon, M.: Studies on the Mechanisms of Metastasis. The Distribution of Arterial Emboli, Cancer Res. **11**:648, 1951.
27. Cooper, R. L., Gilbert, J. A. S., and Lindsay, A. J.: Polycyclic Hydrocarbons in Cigarette Smoke: The Contribution Made by the Paper, Brit. J. Cancer **9**:442, 1955.
28. Cooper, R. L., and Lindsay, A. J.: 3:4-Benzpyrene and Other Polycyclic Hydrocarbons in Cigarette Smoke, Brit. J. Cancer **9**:304, 1955.
29. Cottral, G. E.: Avian Lymphomatosis, Another Egg-Borne Disease, Proc. 53rd Ann. Meeting U. S. Livestock Sanitary Assn. 183, 1950.

VOL. 4 195

30. Crabtree, H. G.: Some Effects of Aromatic Hydrocarbons on Sulfur Metabolism and Tumor Induction in Mice, Cancer Res. **6**:553, 1946.
31. Delarue, N. C.: The Problem of Mammary Carcinoma, Canad. M. A. J. **70**:132, 1954.
32. deLong, R. P., Coman, D. R., and Zeidman, I.: The Significance of Low Calcium and High Potassium Content in Neoplastic Tissue, Cancer **3**:718, 1950.
33. Demerec, M., Moser, H., Hemmerley, J., Blomstrand, I., Demerec, Z. E., Fitzgerald, P. L., Glover, S. W., Hanson, J. F., Nielsen, F. J., and Yura, T.: Bacterial Genetics I., Carnegie Inst. Wash. Year Book **52**:225, 1953.
34. Demerec, M., Witkin, E. M., Labrum, E. L., Galinsky, I., Hanson, J. F., Monsees, H., and Fetherston, T. H.: Bacterial Genetics, Carnegie Inst. Wash. Year Book **51**:193, 1952.
35. Dmochowski, L.: The Milk Agent in the Origin of Mammary Tumors in Mice, Adv. Cancer Res. **1**:103, 1953.
36. Doll, R., and Hill, A. E.: A Study of the Etiology of Carcinoma of the Lung, Brit. M. J. **11**:1271, 1952.
37. Dorn, H.: Morbidity and Mortality From Cancer of the Lung in the United States, in Clemmesen, J., Ed., Cancer of the Lung, Louvain, Council of International Organizations of Medical Sciences, 1953, p. 126.
38. Dorn, H.: The Increase in Cancer of the Lung, Indust. Med. **23**:253, 1954.
39. Dubos, R. J.: Biochemical Determinants of Microbial Diseases, Cambridge, Mass., 1954, Harvard University Press, p. 62.
40. Enticknap, J. B.: Analysis of 1000 Cases of Cancer With Special Reference to Metastasis, Guy's Hosp. Rep. **101**:273, 1952.
41. Ephrussi-Taylor, H.: Current Status of Bacterial Transformations, Adv. Virus Res. **3**:275, 1955.
42. Essenberg, J. M.: Cigarette Smoke and the Incidence of Primary Neoplasm of the Lung in Albino Mice, Science **116**:561, 1952.
43. Ewing, J.: Neoplastic Diseases, Philadelphia, 1940, W. B. Saunders Company.
44. Foote, F. W., Jr., and Stewart, F. W.: The Anatomical Distribution of Intraepithelial Epidermoid Carcinomas of the Cervix, Cancer **1**:431, 1948.
45. Fraenkel-Conrat, H., and Williams, R. C.: Reconstitution of Active Tobacco Mosaic Virus From Its Active Protein and Nucleic Acid Components, Proc. Nat. Acad. Sc. **41**:690, 1955.
46. Furth, J.: Conditioned and Autonomous Neoplasms: A Review, Cancer Res. **13**:477, 1953.
47. Gellhorn, A., and Hirschberg, E.: Investigation of Diverse Systems for Cancer Chemotherapy Screening, Cancer Res. Supp. No. 3, 1955.
48. Goldin, A., Venditti, J. M., Humphreys, S. R., Dennis, D., and Mantel, N.: Studies on the Management of Mouse Leukemia (L 1210) With Antagonists of Folic Acid, Cancer Res. **15**:742, 1955.
49. Goldin, A., Venditti, J. M., Humphreys, S. R., Dennis, D., Mantel, N., and Greenhouse, S. W.: Studies on Toxicity and Antileukemic Action of 6-Mercaptopurine in Mice, Ann. New York Acad. Sc. **60**:251, 1954.
50. Goldthwait, D. A., and Bendich, A.: Effects of a Folic Acid Antagonist on Nucleic Acid Metabolism, J. Biol. Chem. **196**:841, 1952.
51. Greene, H. S. N.: A Conception of Tumor Autonomy Based on Transplantation Studies: a Review, Cancer Res. **11**:899, 1951.
52. Greene, H. S. N.: The Significance of Heterologous Transplantability of Human Cancer, Cancer **5**:24, 1952.
53. Greene, H. S. N.: Heterologous Transplantation of Human Neural Tumors, Cancer **6**:100, 1953.
54. Greene, R. R., Peckham, B. M., Chung, J. T., Bayly, M. A., Benaron, H. B. W., Carrow, L. A., and Gardner, G. H.: Preinvasive Carcinoma of the Cervix During Pregnancy, Surg., Gynec. & Obst. **96**:71, 1953.
55. Greenspan, E. M., Colsky, J., Schoenbach, E. B., and Shear, M. J.: Response of Patients With Advanced Neoplasms to the Intravenous Administration of Alpha-Peltatin, J. Nat. Cancer Inst. **14**:257, 1954.
56. Greenspan, E. M., Leiter, J., and Shear, M. J.: Effect of Alpha-Peltatin, Beta-Peltatin and Podophyllotoxin and Lymphomas and Other Transplanted Tumors, J. Nat. Cancer Inst. **10**:1295, 1950.
57. Greenstein, J. P.: Biochemistry of Cancer, 2nd Ed., New York, 1954, Academic Press, Inc.
58. Gross, L.: The "Vertical" Transmission of Mouse Mammary Carcinoma and Chicken Leukemia; Its Possible Implication for Human Pathology, Cancer **4**:626, 1951.
59. Gross, L.: "Spontaneous" Leukemia Developing in C3H Mice Following Inoculation in Infancy With AK-Leukemia Extracts, or AK-Embryos, Proc. Soc. Exper. Biol. & Med. **76**:27, 1951.
60. Gross, L.: A Filterable Agent, Recovered From AK Leukemic Extracts, Causing Salivary Gland Carcinomas in C3H Mice, Proc. Soc. Exper. Biol. & Med. **83**:414, 1953.
61. Gross, L.: Neck Tumors, or Leukemia, Developing in Adult C3H Mice Following Inoculation, in Early Infancy, With Filtered (Berkfeld N) or Centrifugated (144,000 × 9) AK-Leukemic Extracts, Cancer **6**:948, 1953.
62. Gross, L.: Is Leukemia Caused by a Transmissible Virus? J. Hemat. **9**:557, 1954.

63. Gross, L.: Transmission of AK Leukemic Agent Into Newborn Mice of the C57 Brown/cd Inbred Line, Proc. Soc. Exper. Biol. & Med. **86**:734, 1954.
64. Gross, L.: Induction of Parotid Carcinomas and/or Subcutaneous Sarcomas in C3H Mice With Normal C3H Organ Extracts, Proc. Soc. Exper. Biol. & Med. **88**:362, 1955.
65. Gross, L.: Difference in Susceptibility to AK-Leukemic Agent Between Two Substrains of Mice of C3H Line, Proc. Soc. Exper. Biol. & Med. **88**:64, 1955.
66. Hammond, E. C.: Lung Cancer and Common Inhalants, Cancer **7**:1100, 1954.
67. Hammond, E. C., and Horn, D.: The Relationship Between Human Smoking Habits and Death Rate, J.A.M.A. **155**:1316, 1954.
68. Hamperl, H., Kaufmann, C., and Ober, K. G.: Das Problem der Malignität unter besonder Berücksichtigung des Carcinoma in situ an der Cervix Uteri, Klin. Wchnschr. **32**:825, 1954.
69. Henry, S. A.: Occupational Cutaneous Cancer Attributable to Certain Chemicals in Industry, Brit. M. Bull. **4**:389, 1947.
70. Heston, W. E., and Deringer, M. K.: Test for a Maternal Influence in the Development of Mammary Gland Tumors in Agent-Free Strain C3H Mice, J. Nat. Cancer Inst. **13**:167, 1952.
71. Heston, W. E., Deringer, M. K., Hughes, I. R., and Cornfield, J.: Interrelation of Specific Genes, Body Weight, and Development of Tumors in Mice, J. Nat. Cancer Inst. **12**:1141, 1952.
72. Heston, W. E., and Dunn, T. B.: Tumor Development in Susceptible Strain A and Resistant Strain L Lung Transplants in LA-F_1 Hosts, J. Nat. Cancer Inst. **11**:1057, 1951.
73. Hitchings, G. H., and Elion, G. B.: The Chemistry and Biochemistry of Purine Analogs, Ann. New York Acad. Sc. **60**:195, 1954.
74. Horowitz, N. H., and Leopold, U.: Some Recent Studies Bearing on the One Gene-One Enzyme Hypothesis, Cold Spring Harbor Symp., Quant. Biol. **16**:65, 1951.
75. Howard, L., Erickson, C. C., and Stoddard, L. D.: A Study of the Incidence and Histogenesis of Endocervical Metaplasia and Intraepithelial Carcinoma: Observations on 400 Uteri Removed for Non-cervical Disease, Cancer **4**:1210, 1951.
76. Hueper, W. C.: Environmental Lung Cancer, Indust. Med. **23**:463, 1954.
77. Hueper, W. C.: Recent Developments in Environmental Cancer, Arch. Path. **58**:360, 475, 645, 1954.
78. Kennaway, E. L., and Kennaway, N. M.: Studies in the Incidence of Cancer of the Lung and Larynx, Brit. J. Cancer **5**:153, 1951.
79. Kidd, J. G., and Rous, P.: Carcinogenic Effect of Papilloma Virus on Tarred Skin of Rabbits; Major Factors Determining Phenomenon: Manifold Effects of Tarring, J. Exper. Med. **68**:529, 1938.
80. Kotin, P., Falk, H. L., Mader, P., and Thomas, M.: Aromatic Hydrocarbons. I. Presence in the Los Angeles Atmosphere and the Carcinogenicity of Atmospheric Extracts, Arch. Indust. Hyg. **9**:153, 1954.
81. Kotin, P., Falk, H. L., and Thomas, M.: Aromatic Hydrocarbons. II. Presence in the Particulate Phase of Gasoline-Engine Exhausts, and the Carcinogenicity of Exhaust Extracts, Arch. Indust. Hyg. **9**:164, 1954.
82. Krebs, H. A.: Energy Production in Animal Tissues and in Microorganisms, in Racker, E., Cellular Metabolism and Infections, New York, 1954, Academic Press, Inc., p. 43.
83. Law, L. W.: Origin of the Resistance of Leukemic Cells to Folic Acid Antagonists, Nature **169**:628, 1952.
84. Law, L. W.: Recent Advances in Experimental Leukemic Research, Cancer Res. **14**:695, 1954.
85. Law, L. W., and Boyle, P. J.: Development of Resistance to Folic Acid Antagonists in a Transplantable Lymphoid Leukemia, Proc. Soc. Exper. Biol. & Med. **74**:599, 1950.
86. Law, L. W., Dunn, T. B., and Boyle, P. J.: Neoplasms in the C3H Strain and in F_1 Hybrid Mice of Two Crosses Following Introduction of Extracts and Filtrates of Leukemic Tissues, J. Nat. Cancer Inst. **16**:495, 1955.
87. Leiter, J., Downing, V., Hartwell, J. L., and Shear, M. J.: Damage Induced in Sarcoma 37 With Podophyllin, Podophyllotoxin, Alpha-Peltatin and Quercetin, J. Nat. Cancer Inst. **10**:1273, 1950.
88. Leiter, J., Downing, V., Hartwell, J. L., and Shear, M. J.: Damage Induced in Sarcoma 37 With Chemical Agents. II. Trivalent and Pentavalent Arsenicals, J. Nat. Cancer Inst. **13**:365, 1952.
89. Leiter, J., Downing, V., Hartwell, J. L., and Shear, M. J.: Damage Induced in Sarcoma 37 With Chemical Agents. III. Colchicine Derivatives Related to Trimethylcolchicinic Acid and to Colchinol, J. Nat. Cancer Inst. **13**:379, 1952.
90. Leiter, J., Hartwell, J. L., Kahler, J. S., Kline, I., and Shear, M. J.: Damage Induced in Sarcoma 37 With Chemical Agents. VI. Biphenyl, Fluorene, Phenanthrene and Tropolone Derivatives, J. Nat. Cancer Inst. **14**:365, 1953.
91. Leiter, J., Hartwell, J. L., Kline, I., Nadkarni, M. V., and Shear, M. J.: Damage Induced in Sarcoma 37 With Chemical Agents. IV. Derivatives of Colchiceinamide, J. Nat. Cancer Inst. **13**:731, 1952.
92. Leiter, J., Hartwell, J. L., Kline, I., and Shear, M. J.: Damage Induced in Sarcoma 37 With Chemical Agents. VII. Diphenylpropylamines, Diphenyl Propylenediamines, Diphenylalkanes, and Related Compounds, J. Nat. Cancer Inst. **14**:645, 1953.

VOL 4 195

93. Leiter, J., Hartwell, J. L., Ullyot, G. E., and Shear, M. J.: Damage Induced in Sarcoma 37 With Chemical Agents. V. Derivatives of Colchicine and Isocolchicine, J. Nat. Cancer Inst. **13**:1201, 1953.
94. Leiter, J., and Shear, M. J.: Production of Tumors in Mice With Tars From City Air Dusts, J. Nat. Cancer Inst. **3**:167, 1942.
95. Leiter, J., Shimkin, M. B., and Shear, M. J.: Production of Subcutaneous Sarcomas in Mice With Tars Extracted From Atmospheric Dusts, J. Nat. Cancer Inst. **3**:155, 1942.
96. Levin, M. L., Goldstein, H., and Gerhardt, P. R.: Cancer and Tobacco Smoking, J. A. M. A. **143**:336, 1950.
97. Lorenz, E., Stewart, H. L., Daniel, J. H., and Nelson, C. L.: The Effects of Breathing Tobacco Smoke on Strain A Mice, Cancer Res. **3**:123, 1943.
98. Machle, W., and Gregorius, F.: Cancer of the Respiratory System in the United States Chromate Producing Industry, Pub. Health Rep. **63**:1114, 1948.
99. McKinnon, N. E.: Cancer of the Breast; the Invalid Evidence for Faith in Early Treatment, Canad. J. Pub. Health **42**:218, 1951.
100. McWhirter, R.: Treatment of Cancer of Breast by Simple Mastectomy and Roentgenotherapy, Arch. Surg. **59**:830, 1949.
101. McWhirter, R.: Simple Mastectomy and Radiotherapy in the Treatment of Breast Cancer, Brit. J. Radiol. **28**:128, 1955.
102. Melick, W. F., Escue, H. M., Naryka, J. J., Mezera, R. A., and Wheeler, E. P.: The First Reported Cases of Human Bladder Tumors Due to a New Carcinogen-Xenylamine, J. Urol. **74**:760, 1955.
103. Merewether, E. R.: Ann. Rep. Chief Insp. Factories and Workshops for the Year 1948, London, 1949, Her Majesty's Stat. Off.
104. Miller, E. C., and Miller, J. A.: *In Vivo* Combinations Between Carcinogens and Tissue Constituents and Their Possible Role in Carcinogenesis, Cancer Res. **12**:547, 1952.
105. Miller, J. A., and Miller, E. C.: The Carcinogenic Aminoazo Dyes, Adv. Cancer Res. **1**:340, 1953.
106. Moore, O. S., Jr., and Foote, F. W., Jr.: Relatively Favorable Prognosis of Medullary Carcinoma of the Breast, Cancer **2**:635, 1949.
107. Morris, H. P.: The Experimental Development and Metabolism of Thyroid Gland Tumors, Adv. Cancer Res. **3**:52, 1955.
108. Morris, H. P., Dalton, A. J., and Greene, C. D.: Malignant Thyroid Tumors Occurring in the Mouse After Prolonged Hormonal Imbalance During the Ingestion of Thiouracil, J. Clin. Endocrinol. **11**:1281, 1951.
109. Morton, J. J.: The Etiology of Cancer in the Light of Our Present Knowledge, Ann. Surg. **106**:539, 1937.
110. Müller, F. M.: Tabakmissbrauch u. Lungencarcinom, Ztschr. Krebsforsch. **49**:57, 1939.
111. Oberling, C., and Guérin, M.: The Role of Viruses in the Production of Cancer, Adv. Cancer Res. **2**:253, 1954.
112. O'Brien, F. W., and O'Brien, F. W., Jr.: Surgery and Radiation in Cancer of the Female Breast, Radiology **63**:192, 1954.
113. Papanicolaou, G. H., Traut, H. F., and Marchetti, A. A.: The Epithelia of Woman's Reproductive Organs; a Correlative Study of Cyclic Changes, New York, 1948, Commonwealth Fund.
114. Park, W. W., and Lees, J. C.: The Absolute Curability of Cancer of the Breast, Surg., Gynec. & Obst. **93**:129, 1951.
115. Peacock, P. R., Beck, S., and Anderson, W.: The Influence of Solvents on Tissue Response to Carcinogenic Hydrocarbons, Brit. J. Cancer **3**:296, 1949.
116. Pearson, O. H., Eliel, L. P., Rawson, R. W., Dobriner, K., and Rhoads, C. P.: ACTH- and Cortisone-Induced Regression of Lymphoid Tumors in Man, Cancer **2**:943, 1949.
117. Peller, S.: Cancer in Man, New York, 1952, International Universities Press, p. 337.
118. Pott, P.: Chirurgical Observations, London, 1775, L. Hawes, W. Clark, and K. Collins.
119. Prehn, R. T.: Tumors and Hyperplastic Nodules in Transplanted Mammary Glands, J. Nat. Cancer Inst. **13**:°59, 1953.
120. Reagan, J. W., and Hicks, D. J.: A Study of In Situ and Squamous Cell Cancer in the Uterine Cervix, Cancer **6**:1200, 1953.
121. Reimann, S. P., and Nishimura, E. T.: Attempts at the Chemotherapy of Cancer, J. Michigan M. Soc. **48**:453, 1949.
122. Rhoads, C. P.: Rational Cancer Chemotherapy, Science **119**:77, 1954.
123. Rhoads, C. P.: The Seeing Eye in Science, in Rhoads, C. P., Antimetabolites and Cancer, Washington, 1955, American Assn. for the Advancement of Science, pp. 89-105.
124. Rous, P., and Kidd, J. G.: Carcinogenic Effect of Papilloma Virus on Tarred Skin of Rabbits; Description of Phenomenon, J. Exper. Med. **68**:399, 1938.
125. Sadowsky, D. A., Gilliam, A. G., and Cornfield, J.: The Statistical Association Between Smoking and Carcinoma of the Lung, J. Nat. Cancer Inst. **13**:1237, 1953.
126. Sall, R. D., and Shear, M. J.: Studies in Carcinogenesis. XII. Effect of the Basic Fraction of Creosote Oil on the Production of Tumors in Mice by Chemical Carcinogens, J. Nat. Cancer Inst. **1**:45, 1941.

OL.
4
956

127. Schrek, R., Baker, L. A., Ballard, G., and Dolgoff, S.: Tobacco Smoking as an Etiologic Factor in Disease. I. Cancer, Cancer Res. **10**:49, 1950.
128. Shapiro, J. R., and Kirschbaum, A.: Intrinsic Tissue Response to Induction of Pulmonary Tumors, Cancer Res. **11**:644, 1951.
129. Shear, M. J.: Chemical Treatment of Tumors. IX. Reactions of Mice With Primary Subcutaneous Tumors to Injection of a Hemorrhage-Producing Bacterial Polysaccharide, J. Nat. Cancer Inst. **4**:461, 1944.
130. Shimkin, M. B., Lucia, E. L., Stone, R. S., and Bell, H. G.: Cancer of the Breast. Analysis of Frequency, Distribution and Mortality at the University of California Hospital, 1918-1947, Inclusive, Surg., Gynec. & Obst. **94**:645, 1952.
131. Shope, R. E.: Infectious Papillomatosis of Rabbits, With Note on Histopathology, J. Exper. Med. **58**:607, 1933.
132. Skipper, H. E.: On the Mechanism of Action of Certain Temporary Anticancer Agents: A Review, Cancer Res. **13**:545, 1953.
133. Skipper, H. E., Mitchell J. H., Jr., and Bennett, L. L., Jr.: Inhibition of Nucleic Acid Synthesis by Folic Acid Antagonists, Cancer Res. **10**:510, 1950.
134. Skipper, H. E., Thompson, J. R., and Bell, M.: Attempts at Dual Blocking of Biochemical Events in Cancer Chemotherapy, Cancer Res. **14**:503, 1954.
135. Smith, R. R., and Hilberg, A. W.: Cancer-Cell Seeding of Operative Wounds, J. Nat. Cancer Inst. **16**:645, 1955.
136. Sorof, S., and Cohen, P.: Electrophoretic and Ultracentrifugal Studies on the Soluble Proteins of Various Tumors and of Livers From Rats Fed 4-Dimethylaminoazobenzene, Cancer Res. **11**:376, 1951.
137. Sorof, S., Cohen, P. P., Miller, E. C., and Miller, J. A.: Electrophoretic Studies on the Soluble Proteins From Livers of Rats Fed Aminoazo Dyes, Cancer Res. **11**:383, 1951.
138. Steiner, P. E.: Emphasis in Cancer Research, J. Nat. Cancer Inst. **14**:1205, 1954.
139. Steiner, P. E., Mamion, S. N., Palmer, W. L., and Kirsner, J. B.: Gastric Cancer: Morphological Factors in Five-Year Survival After Gastrectomy, Am. J. Path. **24**:947, 1948.
140. Stewart, S. E.: Neoplasms in Mice Inoculated With Cell-Free Extracts or Filtrates of Leukemic Mouse Tissue. I. Neoplasms of the Parotid and Adrenal Glands, J. Nat. Cancer Inst. **15**:1391, 1955.
141. Stewart, S. E.: Neoplasms in Mice Inoculated With Cell-Free Extracts or Filtrates of Leukemic Mouse Tissues. II. Leukemia in Hybrid Mice Produced by Cell-Free Filtrates, J. Nat. Cancer Inst. **16**:41, 1955.
142. Stock, C. C.: Experimental Cancer Chemotherapy, Adv. Cancer Res. **2**:425, 1954.
143. Stocks, P.: Epidemiology of Cancer of the Lung in England and Wales, Brit. J. Cancer **6**:99, 1952.
144. Sugarbaker, E. D.: The Organ Selectivity of Experimentally Induced Metastases in Rats, Cancer **5**:606, 1952.
145. Taylor, G. W. (Moderator): Significance of Statistical Analysis of End Results in the Treatment of Breast Cancer, Proc. Second Nat. Cancer Conf., New York, 1954, American Cancer Society, Inc., Vol. I, p. 107.
146. Waller, R. E.: The Benzpyrene Content of Town Air, Brit. J. Cancer **6**:8, 1952.
147. Watanabe, S.: The Metastasizability of Tumor Cells, Cancer **7**:215, 1954.
148. Waters, N. F., and Bywaters, J.: Influence of Age of Chickens at Contact Exposure on Incidence of Lymphomatosis, Poultry Sci. **28**:254, 1949.
149. Watson, A. F.: Tar Cancer in Mice. III. An Apparatus for the Preparation Under Standard Conditions of a Highly Potent Carcinogenic Agent of Low Toxicity, Brit. J. Exper. Path. **12**:441, 1931.
150. Weil, C. S., Smyth, H. F., and Nale, T. W.: Quest for a Suspected Industrial Carcinogen, Arch. Indust. Hyg. **5**:535, 1952.
151. Weinhouse, S.: Oxidative Metabolism of Neoplastic Tissues, Adv. Cancer Res. **3**:270, 1955.
152. Welch, A. D., and Nichol, C. A.: Water Soluble Vitamins Concerned With One- and Two-Carbon Intermediates, Ann. Rev. Biochem. **21**:633, 1952.
153. Woglom, W.: Experimental Tar Cancer, Arch. Path. **2**:533, 709, 1926.
154. Wynder, E. L., and Graham, E. A.: Tobacco Smoking as a Possible Etiologic Factor in Bronchiogenic Carcinoma, J. A. M. A. **143**:329, 1950.
155. Wynder, E. L., Graham, E. A., and Croninger, A. B.: Experimental Production of Carcinomas With Cigarette Tar, Cancer Res. **13**:855, 1953.
156. Young, J. S., and Griffith, H. D.: The Dynamics of Parenchymatous Embolism in Relation to the Dissemination of Malignant Tumors, J. Path. & Bact. **62**:293, 1950.
157. Young, J. S., Lumsden, C. E., and Stalker, A. L.: The Significance of "Tissue Pressure" of Normal Testicular and of Neoplastic (Brown-Pearce Carcinoma) Tissue in the Rabbit, J. Path. & Bact. **62**:313, 1950.
158. Zeidman, I., and Buss, J. M.: Transpulmonary Passage of Tumor Cell Emboli, Cancer Res. **12**:731, 1952.
159. Zeidman, I., McCutcheon, M., and Coman, D. R.: Factors Affecting the Number of Tumor Metastases. Experiments With a Transplantable Mouse Tumor, Cancer Res. **10**:351 1950.

Editorials

SEX HORMONE DEFICIENCY OF ADRENAL ORIGIN IN OLDER MEN AND WOMEN

The Climacteric as Adrenopause

VOL.
4
1956

Intelligent older patients in the doctor's office these days expect to be studied for diseases that they themselves cannot detect. They wish to be protected against asymptomatic lesions and disorders. It appears to me that adrenocortical deficiency in the production of androgens and estrogens is one of the common causes of subclinical disease in older people, a conclusion which may be drawn from the following working hypothesis: osteoporosis is a common disease after 50 years of age[1]; senile osteoporosis responds to treatment for so-called postmenopausal osteoporosis and is "really the same disease"[2] estrogens and androgens are effective therapy for this condition, indicating that it is due to their deficiency; estrogens and androgens are produced by the adrenal cortex after 50 years of age[3]; hence this form of adrenocortical insufficiency is one of the common causes of disease in older people. The syllogism refers to osteoporosis and the same line of reasoning may also explain atherosclerosis as the result of relative adrenocortical insufficiency in the production of estrogen. For example, estrogens have been demonstrated to prevent experimental atherosclerosis.[4] Such recent studies have "lent support to the concept that estrogens are decisively responsible for the relative immunity of premenopausal women to coronary disease"[4] and have encouraged the trial of estrogen therapy to prevent recurrences of myocardial infarction in men. Certainly estrogen deficiency in the male must be attributed to relative adrenocortical, not gonadal, deficiency. A study in progress of postmenopausal women has indicated that in those patients who have had coronary thrombosis, the estrogen excretion is lower than in those who have not, thus indicating that estrogen deficiency after the ovaripause is sometimes associated with atherosclerosis.[5] Clinicians have long associated a number of systemic diseases with the menopause. They blamed the hormonal changes that occurred for endocrine diseases, such as goiter, hyperthyroidism, and acromegaly; for metabolic disorders such as obesity and diabetes; for diseases of the skeleton, such as rheumatoid arthritis and osteoporosis; and for vascular diseases such as hypertension, arteriosclerosis, and, now, coronary thrombosis. Current biochemical research may show that such diseases are the result of abnormal adrenal production in later life.

Adrenal sex steroid insufficiency has not been recognized as a clinical entity for several reasons: the metabolic effects are subclinical; disorders of the menopause have been attributed to a readjustment of the nervous system; and since

cessation of ovarian function is normal, it was thought that no abnormal deficiency state need be expected. Those forms of adrenal insufficiency characterized by a diminished or disproportionate secretion of sex steroids are as yet only in part defined. Well-known adrenocortical syndromes are produced by lowered secretion (Addison's disease); by increased secretion of glucocorticoids (Cushing's syndrome); and by excess androgens (adrenogenital syndrome) or, rarely, by an excess of estrogens. Recently a syndrome caused by excess mineralocorticoids (aldosteronism) has been described. And the possibility of a disease resulting from a lack of mineralocorticoids has been suggested. Osteoporosis results when the adrenal cortex fails to produce an adequate amount of the anabolic sex steroid hormones, but continues to produce effective amounts of glucocorticoids. Such a disorder may cause trouble in the cardiovascular and other systems, as well as in the skeleton.

Before the present era of endocrine and biochemical knowledge, the disorders of the menopausal period in women, and even some diseases of older men, were usually treated with sedatives and psychotherapy. This practice was based on several misapprehensions, as follows: female sex hormone is not produced after the menopause and is obviously no longer needed; the administration of estrogens will prolong the transitional period and delay readjustment to the climacteric; continuation of periodic endometrial bleeding will occur; and estrogens are not normally maintained in older people and may therefore have a carcinogenic effect on the uterus and breast. Hence, the patient had better take a bromide and tolerate the symptoms.

But these clinical opinions may be reconsidered as follows: First, all women experience the menopause, but many of them have no change in health, vasomotor symptoms, vaginal atresia, or systemic disease. They have no "change of life." Second, normal women produce significant amounts of estrogen after the menopause (Table I). Third, the adrenal cortex is a normal source of androgens, estrogens, and progesterones. Finally, the general metabolic function of the sex hormones throughout the body, although poorly defined, has been demonstrated.

The adrenal production of androgens and estrogens is minimal before puberty, increasing with the pubescent activity of the pituitary, persisting after the end of gonadal function in women, and continuing throughout the later life of both sexes. In the male, two-thirds of the total androgen secretion is at all times derived from the adrenal cortex, not from the gonad. Estrogen, which is produced steadily to meet metabolic requirements, comes entirely from the adrenal cortex in the male. In women the adrenal is the only source of the necessary androgen.

There is evidence that estrogen and androgen production continues after 50 years of age in both sexes and is sustained at a significant level for the remaining decades of life. A group of investigators at the Worcester Foundation for Experimental Biology has determined the excretion of neutral steroids in groups of men and women from 20 to 90 years of age, divided by decades into groups of about fifty subjects. Androgens were determined chemically and estrogens by bioassay. Division of the twenty-four hours was made according to waking, day, and sleeping periods.

These data—the most recent, extensive, and reliable—allow an evaluation of sex steroid metabolism in adult men and women and are especially valuable for the interpretation of endocrine function in the last four decades of life. Tables I and II were selected from the extensive data tabulated and analyzed in the article.[6]

TABLE I. THE MEAN OUTPUT OF VARIOUS URINARY STEROIDS IN FEMALE SUBJECTS

DECADE	17 KETOSTEROIDS MG./HR. (DURING DAYTIME)	ESTROGEN R.U./24 HR.
20–29	0.363	33.6
30–39	0.355	34.39
40–49	0.270	18.75
50–59	0.234	10.73
60–69	0.189	6.64
70–79	0.174	7.16
80–89	0.121	6.88

From Pincus, G., Romanoff, L. P., and Carlo, J.: The Excretion of Urinary Steroids by Men and Women of Various Ages. J. Gerontol. **9**:113, 1954.
*Mean values of groups numbering in each decade from 27 to 67 women.

In the eighth decade the androgen excreted as 17-ketosteroid by the female subjects (Table I) was still 50 per cent of that measured in the prime of life; estrogen excretion is maintained at a significant level, even at 90 years of age. The daily estrogen excretion, from 10.73 to 6.64 rat units per 24 hours, was about two-thirds to three-fourths of that found in males in every decade from 20 to 90 years of age. Since estrogen in the male functions in systemic, e.g., extragenital metabolism, the amount is an indication of that required for these metabolic processes. It is reasonable to assume that the postmenopausal amount found in women is enough for these purposes and is *normally required for them*. Chemical determinations in oophorectomized and adrenalectomized patients prove that the excretion of estrogen is abolished after the adrenal has been removed.[3]

Many women over 50 years of age have no symptoms of sex hormone climacteric; in others the dysfunctions are transient; but some women have outspoken complaints, which might properly be attributed to postmenopausal adrenal failure, or, as Albright called it, the "adrenopause."

The diagnosis of chronic adrenocortical sex hormone insufficiency should be considered if the autonomic postmenopausal nervous symptoms fail to clear up after a few months. True, there is no demonstration that the spontaneous disappearance of these reactions indicates a revival of sex hormone secretion, since the symptoms might stop because of hypothalamic readjustment to the deficiency state. Nevertheless, the persistence of vasomotor and hyperhydrotic disorders for more than a few months is a practical indication for diagnostic investigation. Some women may have these distressing signs for five, ten, or even twenty years after the menopause. A rapid development of unhealthy bodily changes, such as obesity, hypertension, osteoporosis, arthritis, coronary insufficiency, or thrombosis; and complaints of exhaustion, weakness, fatigue, and depression, or local symptoms such as atresia, vaginitis, urethritis, and cystitis, warrant study for endocrine deficiency. Laboratory screening pro-

cedures that may be useful are: histologic interpretation of vaginal epithelial smears for estrogen effect; bioassay of the amount of pituitary gonadotropin and estrogen in the urine; and chemical determination of the excretion of androgens. Measurement of circulating thyroxine, as serum protein-bound iodine, may also be of value. In the absence of laboratory facilities, a therapeutic trial, accomplished, if possible, without psychologic prejudice, is indicated. The laboratory investigation may seem troublesome and expensive for a subclinical condition, but if it discloses a metabolic disorder that is contributory to systemic degeneration, productive of structural breakdown, and conducive to senile decrepitude, it is worth the time and expense, especially as effective hormone replacement medication is available.

Oral therapy with an estrogen/androgen combination of steroids is necessary when a chronic deficiency state has been established. The proper dosage of these hormones has not been determined by critical, long-term metabolic and pathologic studies. At present, the maximum daily oral dose that is tolerated by the individual patient is advised. The daily amount of estrogen and androgen, and the ratio between the two, must be titrated in each patient. Usually one part of estrogen will neutralize the masculinizing action of twenty parts of androgen. The patient should be instructed and persuaded that this medication replaces substances that were normally present in her tissues, in amounts similar to those natural for her age, without which systemic disorders may occur; she should be told that there is a glandular deficiency, not of the ovary, which has finished its work, but of the adrenal gland, which should be carrying it on; that sometimes overdosage, or interrupted dosage leads to uterine bleeding; and that the hormones are not carcinogenic. She should be told that a complete physical examination, semiannually, is advisable.

TABLE II. MEAN OUTPUT OF VARIOUS URINARY STEROIDS IN MALE SUBJECTS

DECADE	17 KETOSTEROIDS MG./HR. (DURING DAYTIME)	ESTROGEN R.U./24 HR.
20–29	0.524	11.50
30–39	0.458	12.44
40–49	0.430	12.01
50–59	0.328	10.64
60–69	0.318	9.61
70–79	0.222	12.48
80–89	0.181	10.48

From Pincus, G., Romanoff, L. P., and Carlo, J.: The Excretion of Urinary Steroids by Men and Women of Various Ages, *J. Gerontol.* **9**:113, 1954.

In men there is no regular (i.e., occurring in every man) degeneration of the gametogenic tissue, and therefore no spermatopause. Neither is there any evidence of a hormonal climacteric—an abrupt change in the amount of hormone secreted from one category of magnitude to another at 50 or 60 years of age. As shown in Table II the transition is gradual from decade to decade and no sudden stepdown occurs.

The times of life under consideration, the sixth or seventh decades, are accompanied by a reduction of androgen excretion of about 23 per cent between the fifth and sixth decades, and 3 per cent reduction from the sixth to the seventh (Table II). The changes are not great enough to justify the concept of a male climacteric.

After 50 years of age, for a possible remaining forty years of life, it is normal for women to produce androgen, represented by an average excretion of about 0.180 mg. of 17-ketosteroid per hour, and estrogens represented by an average excretion of about 7.85 rat units per 24 hours.* During the same period of life it is normal for men to produce androgens, represented by an average excretion of 0.262 mg. of 17-ketosteroids per hour and estrogens represented by an average excretion of about 10.95 rat units per 24 hours.*

There are, however, some older men who have hypoandrogenism, which is certainly not corrected spontaneously by the adrenocortical secretion of androgens. To paraphrase Albright, these men have a "leydigpause and an adrenopause." They may have many symptoms that are similar to those of women with postmenopausal, uncompensated hypo-ovarianism. They have excessive amounts of gonadotrophic hormone in the urine.[7] The pathogenesis of this type of male hormone insufficiency is also adrenocortical failure; and the treatment of hormone deficient men is as necessary and beneficial as the treatment of women who have adrenal hypoestrogenism.

Careful observation of patients with metabolic disorders shows that deficiencies serious enough to cause actual tissue damage may be asymptomatic; the physician should adopt screening procedures to detect such subclinical conditions. These asymptomatic disorders have been well demonstrated in thyroid deficiency and in diabetes, and may well be present in sex steroid adrenocortical insufficiency. A study by Sobel, Hoffman, Marmorston, and Starr[5] showed a deficiency of both estrogen and thyroxine in postmenopausal women who have had coronary thromboses. These deficiencies were found by a careful selection of patients, based on the occurrence of coronary artery thrombosis. In other words, the deficiency was demonstrated after a serious pathologic event. A more intelligent practice among internists would be the use of screening procedures in asymptomatic patients to detect the deficiency that may be leading to a pathologic condition. We must get beyond practicing symptomatic medicine—waiting for the patient with cancer to have lesions, for the patient with tuberculosis to have a hemorrhage, or the patient with postmenopausal estrogen deficiency to have a vertebral collapse from osteoporosis.

Reliable screening procedures for many important diseases are now in use. The serologic test is used to detect syphilis; the chest x-ray is invaluable in the discovery of asymptomatic tuberculosis; urine examination or measurements of fasting blood sugar are useful in indicating diabetes; and the general use of serum protein-bound iodine determination is advisable to detect subclinical hypothyroidism. Screening procedures for adrenal hormone insufficiency, however, are much less well established.

*Arithmetic average of the values in Tables I and II of the last four decades.

The most efficient and economical screening procedure for the practicing physician to use in detecting estrogen-androgen deficiency may turn out to be the determination of the gonadotrophic pituitary hormone excretion, which seems to accompany the fundamental disorder in these states. Albert[8] has described a new means for extracting follicle stimulating hormone (FSH) and has given an excellent bioassay standardization for urinary gonadotrophin, using the increase in adult rat ovarian weight, the end point adoped by Heller and Myers[7] in their study of elderly hypogonad men in 1944. This procedure is applicable to both sexes and would perhaps yield significant findings in patients with sex steroid adrenocortical insufficiency. Direct determination of the 17-ketosteroids in the urine is now a well-established laboratory procedure. A larger clinical demand for this measurement would, we hope, lead to a reduction in the expense for it. Urinary estrogen determinations are not so well established and must await further chemical study.

The mistaken view of the female climacteric arose from the clinical deduction that the end of gametogenesis meant the end of female sex hormone production. A similar fallacy about male physiology arose from the belief that the continuation of sperm production meant undiminished androgen secretion. Neither conclusion is true: the menopause alone does not establish the end of sex steroid secretion nor does the continuation of spermatogenesis exclude the possibility of sex hormone deficiency in men. Estrogen or androgen deficiency in older people should not be classified by the physician as an example of the normal, but should be recognized as a pathologic deficiency state, and corrected.

VOL. 4 1956

REFERENCES

1. Albright, F.: Cushing's Syndrome, Harvey Lecture Series **38**:123, 1942-43.
2. Cooke, A. M.: Osteoporosis, Lancet **1**:877, 929, 1955.
3. Bergenstal, D. M., Huggins, C., and Dao, T.: Metabolic Effects of Adrenalectomy in Man, The Human Adrenal Cortex, Ciba Foundation on Endocrinology, Boston, 1955, Little, Brown & Company, vol. 3, p. 415.
4. Katz, L. N.: The Role of Diet and Hormones in the Prevention of Myocardial Infarction, Ann. Int. Med. **43**:930, 1955.
5. Marmorston, J., Hoffman, O., Sobel, H., and Starr, P.: Unpublished data.
6. Pincus, G., Romanoff, L. P., and Carlo, J.: The Excretion of Urinary Steroids by Men and Women of Various Ages, J. Gerontol. **9**:113, 1954.
7. Heller, C. G., and Myers, G. B.: The Male Climacteric, Its Symptomatology, Diagnosis, and Treatment, J. A. M. A. **126**:472, 1944.
8. Albert, A.: Procedure for Routine Chemical Determination of Urinary Gonadotrophin, Proc. Staff. Meet., Mayo Clin. **30**:552, 1955.

PAUL STARR, M.D.
University of Southern California
Los Angeles, Calif.

USE OF DRUGS IN THE TREATMENT OF PERIPHERAL VASCULAR DISEASE

Introduction.—The veritable shower of drugs which have been developed for dilating peripheral vessels by antiadrenergic action, by performing medical sympathectomy with ganglionic blocking and possibly by action on the central nervous system, has brought renewed interest in the possibilities of medical treatment of hypertension and peripheral vascular disease. Many of these drugs are valuable in both conditions but discussion is here confined to their use in peripheral vascular disease.

Notes on Older Drugs.—During the past three or four decades, there have been a number of drugs ranging from muscle (Kallikrein) and pancreatic (Depropanex) extracts to intravenous infusion of ether used for the treatment of ischemia due to arterial obstruction. Among those that enjoyed greater popularity were papaverine, histamine, hypertonic saline, and nicotinic acid recently revived in the form of Roniacol. Of course, perhaps the most ancient and certainly the most extensively used vasodilator is ethyl alcohol. Recently, it has occasionally been administered intravenously with doubtful clinical benefits although vasodilation following its use could be documented. There were numerous reports of limited success with these drugs in the treatment of peripheral vascular disease; but perusal of the literature leaves one with the impression that none were more than somewhat helpful. The difficulties of trying to compare the effect of a drug on a new group of patients with disease of long standing with a group treated previously by different means are apparent. There is also the fact to be reckoned with that when a new drug causes a ripple of enthusiasm the patients are liable to receive better general care. In short, one is never certain that the new agent under study does much more than cleanliness, careful dressing of ulcers, and physical methods such as reflex dilation by warming the body, positive-negative pressure, and Buerger's exercise regime.

Types of Newer Drugs.—Of the numerous drugs which have been developed recently on perhaps more physiologically specific bases, some at least have appeared to be moderately successful.

The pharmacologic actions of these drugs can be classed under three heads: (a) antiadrenergic (adrenolytic); (b) ganglionic blockade; (c) central (inhibition of the sympathetic centers in the brain).

Several agents appear to exhibit action in more than one site. For various reasons some of the drugs have emerged as more useful than others. But before discussing the individual drugs, a few of the factors or criteria employed in choosing a drug might be mentioned.

Choice of Drug.—(a) In long-drawn-out illness, to be able to administer a drug orally is of first importance because other modes of administration tend to disrupt the individual's way of life. Intramuscular administration is next in

preference because the patient can be taught to do it at home. Intravenous and intra-arterial injection are objectionable in that they require constant attendance by the patient at a clinic or office or constant attendance by the physician in the patient's home. (b) Incommoding hypotensive effects must be avoided. (c) Long duration of action—perhaps the longer the better—is highly desirable. (d) Unpleasant or dangerous side actions must be avoided. (e) Consideration must be given to the major site of action, e.g., muscle or skin.

Individual Drugs.—Let us first take up the drugs generally classed as antiadrenergic (adrenolytic), since it is among this group that the most useful are to be found. Among them are Dibenamine, Dibenzyline, a cousin of Dibenamine, and Ilidar.

Dibenamine (N, N-dibenzyl-β-chloroethylamine), although a potent antiadrenergic drug and vasodilator, has not been used widely because it must be given intravenously and because it has exhibited rather strong side effects—congestion of nose and throat and fixation of the pupil and sometimes marked lowering of blood pressure. Probably the fact that Dibenzyline (N-pheno xyisopropyl-N-benzyl-β-chloroethylamine), a close relative of Dibenamine, can be given by mouth and exhibits most of the actions of Dibenamine, has given rise to its more general use. Most of our experience with antiadrenergic drugs has been drawn from working with this drug. In addition to its antiadrenergic action there is some evidence that the drug may also have a direct action on vascular musculature.

There seems to be little room for doubt that Dibenzyline given either in single intravenous doses or orally over long periods of time increases the blood flow to the extremities in both normal individuals and individuals with marked arteriosclerosis obliterans and Buerger's disease. Vasodilation of the vessels of the extremities can usually be obtained at levels of dosage below those resulting in discommoding postural hypotension. The side effects of adrenergic blocking, i.e., stuffiness of nose and throat and fixation of pupil, are usually present to a mild degree which, although unpleasant, need not be serious. In this connection, it must be noted that this drug is contraindicated in asthmatic patients because of the congestion of the upper respiratory passages which accompany its use. Tachycardia is also a frequent result which may be avoided to some degree by simultaneous administration of reserpine. Another side effect of Dibenzyline was noticed to be failure of ejaculation during sexual intercourse. This is in contrast to the impotence which develops in patients being treated with the ganglionic blocking agents.

One of the interesting points concerning dosage of the drug is the remarkable range of effective dose and tolerance. For example, in Raynaud's disease with relatively little organic arterial disease, 10 mg. a day may suffice to obtain a therapeutic effect and attempts to increase the dose may result in severe postural hypotension. In a very advanced case of obliterative endoarteritis, however, 400 mg. daily was tolerated well and indeed was required to secure vasodilation. It has seemed to us at times that the more extensive the organic vascular disease, the higher the tolerated dose. There are, however, exceptions to this generalization.

VOL 4 195

Considerable success has been obtained in the treatment of gangrenous ulcers which were progressing at such a rate that amputation was being seriously considered. Often relief of pain was dramatically prompt. Of greater importance was the fact that these progressive gangrenous ulcers began to heal slowly. First, demarcation of the border of gangrene became clear, and then healing began. The degree of success in relief of intermittent claudication, if any, was less impressive and very slow. This fact may be the clinical expression of the physiologic observation that the predominant action of Dibenzyline is on the vessels of the skin, perhaps benefiting those of the muscle only secondarily.

More recently a combination of the use of Dibenzyline and local use of enzymes (Varidase) for digestion of necrotic material has been used. It appears that the maximal effect of Dibenzyline can be increased by the use of Varidase and, conversely, that the maximal healing under use of the enzymes can be increased by administration of Dibenzyline. Where the institution of both agents has been simultaneous, healing of gangerous ulcers has seemed to take place more rapidly. A fact which is, however, more important than the speed of healing, is that healing has been completed under the influence of the combined use of these two agents with different but complimentary pharmacologic effect when neither agent alone seemed capable of doing the job.

OL.
4
956

Ilidar (6-allyl-6, 7-dihydro-5-dibenz (c, e) azepine) can also be used over long periods of time and may be given orally. The fact that blood flow to the lower extremities increases has been documented plethysmographically and by surface temperature recordings. Postural hypotension may occur. The drug is contraindicated in patients with asthma or emphysema as bronchospasm may be elicited by the drug.

Naturally the drug is most effective in vasospasm. Its effective use in Buerger's disease and in arteriosclerosis obliterans suggests that in relieving vasospasm and by counteracting the vasoconstrictor effect of the circulating catechol amines over long periods of time, the drug may enhance the development of the collateral circulation. Relief of coldness of the extremities and diminution of intermittent claudication and healing of ischemic ulcers of long standing may follow. It must be emphasized that improvement of the circulation may take place very slowly. One must persist in daily use of the drug for several weeks before deciding that it is useless.

The second group of agents—those whose action is presumably on the ganglia—have not been practical in the treatment of peripheral vascular disease because of the development of hypotension.

Etamon (Tetraethyl ammonium chloride), because it may produce severe postural hypotension, has been useful only as a substitute for paravertebral block in estimating the spastic element in occlusion.

Hexamethonium, on the other hand, has been used to a greater extent, but although good peripheral dilation can be obtained, the accompanying postural hypotension is usually severe enough in most instances to prevent the patient from assuming the erect posture. Thus it could be used effectively in bed patients for the most part.

Ansolysin (1:5 bis-(1-methyl-pyrrolidinium) bitartrate) has been used but little. Essentially the same difficulties arise in its use as in hexamethonium.

Priscoline (2 benzyl-4-5-imidazoline hydrochloride) is generally referred to as a sympatholitic drug and is believed to act at the nerve endings although there is also evidence for antiadrenergic action as well. Priscoline is effective by mouth although it has been used extensively by intra-arterial injection. It is said to act on blood vessels in the muscles as well as those in the skin; but plethysmographic studies in this laboratory have not succeeded in demonstrating an effect on muscle blood flow. It can be used over long periods of time and like Dibenzyline the relief of pain in gangrenous ulcers may be prompt followed by slow healing. Some authors have found it promptly effective in relieving intermittent claudication but doses large enough to be effective also cause burning of the skin and unpleasant sensations of warmth in blush areas. There may occasionally be mild gastrointestinal complaints, but postural hypotension is not a problem.

Hydergine (mixture of dihydroergocornine, dihydroergocristine and dihydroergocryptine) is capable of depressing constrictor tonus by inhibition of the vasomotor centers. It is also said to have a weak adrenolytic action, and has been used in chronic peripheral vascular insufficiency as well as in arterial hypertension. Preparations are available for intramuscular, oral, and sublingual administration. Veratrum preparations are likewise centrally acting vasodilators, but have not been used to any extent in peripheral vascular disease.

On a few occasions we have compared the effects of arterial injection of histamine, Priscoline, and Dibenzyline with those following intravenous injection. It may be true that with Priscoline, dilation of the vessels in the affected extremity by arterial injection may be obtained with fewer systemic effects; but the action following intra-arterial injection is extremely short. That intra-arterial injection is effective in dilating the vessels of an extremity has been made clear. It is, however, an even more exacting procedure than intravenous injection and either route seems unnecessary if similar results can be obtained by oral administration. The disadvantages of having the patient return frequently, of intermittency in effective drug level, and even the unpleasantness of repeated injections outweigh the possible advantages of obtaining somewhat better vasodilation. The results so far reported do not seem to justify its use except perhaps in acute occlusion.

In summary then, the outlook at present for treating occlusive peripheral vascular disease medically looks much brighter than it has in the past.

J. Murray Steele, M.D., and
Walter Redisch, M.D.
New York University Research Service
Goldwater Memorial Hospital
Welfare Island
New York, N. Y.

Erratum

The affiliation of Dr. Karl Harpuder, whose article entitled "Basic Medical Principles in the Treatment of the Chronically Ill Patient" appeared in the August issue of the Journal, pages 170-176, was given incorrectly after his name on the first page of the article. Dr. Harpuder should have been listed as Attending Physician, Department of Physical Medicine and Rehabilitation at Montefiore Hospital, and Assistant Clinical Professor of Medicine, Columbia University.